

基于改进随机森林分类器在 RGBD 面部表情上的应用研究

张御宇¹, 倪蓉蓉², 杨彪¹

(1.常州大学信息科学与工程学院, 江苏 常州 213164)

(2.常州纺织服装职业技术学院, 江苏 常州 213164)

[摘要] 面部表情识别是机器感知人类情绪变化的重要途径. 利用面部 RGB 图像可以进行表情识别, 但是容易受到光照变化影响, 而且较难刻画细微表情变化. 对采用 RGBD 面部图像识别 6 种基本面部表情(高兴、悲伤、愤怒、沮丧、恐惧以及惊讶)进行研究. 首先利用深度图像鲁棒地检测面部; 然后在面部灰度图像中检测并跟踪二维面部标记点, 并添加对应的深度信息构造深度面部几何特征, 从而有效识别细微表情变化; 最后利用基于特征选择的随机森林分类器对不同面部表情进行识别. 基准数据库上的对比实验结果表明本文算法的表情识别准确率高于主流基于手动提取特征的面部表情识别方法, 接近基于卷积神经网络的识别算法性能.

[关键词] 面部检测, 面部表情识别, RGBD 面部图像, 特征选择随机森林, 面部几何特征

[中图分类号] TP391 [文献标志码] A [文章编号] 1001-4616(2019)01-0082-08

Research on Bimodal Facial Expression Based on Improved Random Forest Classifier

Zhang Yuyu¹, Ni Rongrong², Yang Biao¹

(1.School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

(2.Changzhou Vocational Institute of Textile and Garment, Changzhou 213164, China)

Abstract: Facial expression recognition is an important way for machines to understand expression changes of human beings. Expression recognition using RGB facial images suffers from illumination changes and detail depicting in minor emotion changes. This work employ RGBD facial images to recognize 6 basic facial expressions (happiness, sadness, angry, disgust, fear and surprise). Depth images are used for robust face detection initially. Gray-scale facial images are employed to detect and to track 2D facial landmark points from the detected face region. Then, corresponding depth information is added into these points to construct depth facial geometrical feature which can recognize minor expression changes more effectively. Finally, a random forest classifier based on feature selection is designed to recognize different facial expressions. Results of comparative evaluations on benchmarking datasets verify the fact that our approach outperforms several state-of-the-art facial expression approaches, which use hand-crafted features, in recognizing six basic facial expressions. Meanwhile, our approach achieves almost similar performance comparing with the convolutional neural network based expression recognition algorithms.

Key words: face detection, facial expression recognition, RGBD facial images, feature selection based random forest, facial geometrical feature

面部表情识别在揭示人类意图、情感及其他内在状态方面发挥着极大的作用, 是机器感知人类情绪变化并与人沟通的重要手段, 在人机交互、计算机辅助学习、健康监控、辅助驾驶、异常行为检测等方面获得了广泛应用.

收稿日期: 2018-08-16.

基金项目: 国家自然科学基金(61501060)、江苏省科技厅青年基金(BK20150271)、江苏省道路载运工具新技术应用重点实验室开放课题(ZMF15020068).

通讯联系人: 杨彪, 博士, 讲师, 研究方向: 机器视觉、模式识别. E-mail: yb6864171@cczu.edu.cn

面部表情识别主要包括 3 个步骤:面部区域检测、面部特征提取以及面部表情分类^[1]。面部区域检测通常使用级联检测器^[2]实现,对于检测结果,需要进一步利用眼睛、嘴巴等器官的位置进行矫正^[3],从而减少后续特征提取与表情识别中无关信息的干扰;对于检测到的面部区域,需要提取能够反映表情变化的特征,主要有基于几何特征的方法与基于外表的方法:前者通常提取面部典型区域,如嘴、眼睛、眉毛、鼻子等的位置与形状信息^[4],而后者采用面部的整体或局部区域作为特征^[5],近年来,也有学者利用运动特征来识别面部表情^[26];最后,利用特定分类器(支持向量机、随机森林、神经网络等)对面部特征进行训练与分类,从而识别不同的面部表情。

但是,传统面部表情识别算法通常在 RGB 或灰度面部图像下进行,因而在描述细微表情变化时存在局限性。同时,光照变化、异物遮挡也会造成面部检测不准确,进而降低表情识别率。近年来,有学者研究了如何利用 3D(三维)面部图像研究表情识别问题^[6],提取 3D 面部图像的几何特征进行表情分类,准确率较高。同时,基于 2D+3D(二维+三维)的面部表情识别方法也获得了一定关注,其研究重点是对多模态面部图像进行有效融合^[7],尤其是特征级融合。但是,三维图像的获取比较困难(需要昂贵的高分辨率图像传感器进行三维重构),而且计算复杂度偏高,不利于实际应用。

针对上述问题,本文以面部 RGBD 图像作为研究对象,提出了一种基于深度面部几何特征与改进随机森林算法的面部基本表情(高兴、悲伤、愤怒、沮丧、恐惧以及惊讶)识别方法。本文的主要工作及创新点:(1)提出了一种对光照变化鲁棒的面部检测算法。传统 RGB 或灰度图像对光照变化敏感,而深度图像对光照变化有较强鲁棒性,因此本文在深度图像上检测面部区域并映射到 RGB/灰度图像;(2)提出一种深度面部几何特征。本文对传统 2D 面部几何特征进行改进,引入深度信息,构造深度面部几何特征,能够更好地识别细微表情变化;(3)提出一种改进的随机森林分类算法。考虑到并非所有特征都能够为分类器提供有效信息,因此对输入分类器的特征进行筛选,增加表情识别准确率。同时,本文使用 Kinect 2.0 设备搭建面部表情识别系统,采集实际 RGBD 面部表情图像进行算法性能分析。

1 相关工作

鉴于面部表情识别在人机交互领域的重要作用,前期对其开展了大量研究,主要可分为基于 2D 面部图像与基于 3D 面部图像的表情识别。

对于 2D 面部表情识别,主要是在面部 RGB 图像或灰度图像上提取与表情变化相关的特征,并训练分类器识别不同表情,常用的特征包括外表特征、几何特征以及运动特征。面部外表特征包括像素强度^[8]、Gabor 纹理^[9]、局部二值模式(local binary patterns, LBP)^[10]以及方向梯度直方图(histogram of oriented gradients, HoG)^[11]等,它们描述了面部的全局或细节信息,可以在一定程度上反映面部表情。但是,外表特征通常对全部面部进行分析,而面部表情变化通常与眼睛、嘴角、眉毛等关键区域高度相关。因此,有学者提出可用于面部表情识别的几何特征,即检测表情变化关键区域的面部特征点,并通过这些点之间的几何关系来识别不同表情^[12]。随着面部表情识别在视频分析领域的发展,越来越多的学者开始关注描述面部表情变化的运动特征,主要有光流^[13]、运动历史图像(motion history images, MHI)^[14]以及成卷 LBP(Volume LBP)^[15]。

除了上述手动提取的特征外,深度学习网络可以自动提取任务相关的特征并进行分类,并且具有更好的性能。Zhao 等人利用深度置信网络(DBNs)的无监督学习能力自动学习面部特征,并用多层感知机(MLP)进行分类,其分类精度优于主流的支持向量机、最近邻分类器以及稀疏表示分类器等^[16]。Lopes 等人利用卷积神经网络(CNN)识别面部表情,并利用数据增益策略解决 CNN 训练中需要大量样本的问题^[17]。Yu 等人则通过调整集合权重将不同 CNN 组合起来识别面部表情^[18]。但是,深度学习网络对样本数量及硬件要求较高,而且参数调整过程复杂,较难普遍实现。

基于 2D 面部图像的表情识别具有数据易获取、特征提取简单等优点,但是 2D 面部图像中不包含深度信息,因此在识别面部细微表情变化时存在不足,因此也有学者对 3D 面部图像的表情识别进行了研究。部分学者将 2D 面部图像中的表情特征拓展到 3D 面部图像,提出了诸如 3D 面部标记点^[19]、3D 小波纹理^[20]以及 3D 有向光流直方图(histogram of oriented optical flow, HoOF)^[21]等特征进行面部表情识别。也有学者利用 3D 面部图像的特点设计新的特征,如 Zhen 等人提出的肌肉运动模型(muscular movement

model,MMM)特征^[22].

3D 面部图像能够提供丰富的面部表情信息,但是,获取 3D 面部图像的硬件成本较高,而且从 3D 面部提取特征其复杂度也远高于从 2D 面部提取特征,较难普遍推广. 随着 Kniect 等消费级 RGBD 图像传感器的不断发展,研究者能够以较低的成本同时获得目标的 RGB 图像与对应的深度图像,从而改善单纯 RGB 图像无法描述目标深度信息的不足. 目前,已有较多研究使用 Kniect 进行面部表情识别,但通常需要利用获取的面部 RGBD 数据进行 3D 重构,譬如 Zhang 等人通过 Kniect 2.0 获取 3D 重构后面部的点云特征,进而识别悲伤、高兴以及中性 3 种表情^[23]. 总体来说,目前基于 Kniect 的面部表情识别研究需要大计算量的 3D 面部重构,实时性较差.

2 提出方法

2.1 面部表情识别算法流程

如图 1 所示为本文算法流程,主要分为图像预处理、特征提取以及分类器训练. 其中图像预处理部分主要完成图像灰度化、面部检测以及降采样;特征提取部分主要提取深度面部几何特征;分类器训练部分主要利用所提取的特征训练改进的随机森林分类器,并利用该分类器识别面部表情.

2.2 图像预处理

面部表情图像(标准数据库/实际数据集)通常包含较多背景信息,因此需要对图像进行预处理,主要包括以下步骤:

(1)面部检测
传统面部检测算法通常在 RGB 或灰度图像下进行,易受光照变化影响. 深度图像对光照变化有较强鲁棒性,因此本文采用 Viola-Jones 算法^[24]在深度图像上进行面部检测,并将检测得到的面部区域映射到面部灰度图像中. 由于输入的是 RGBD 图像,因此需要将 RGB 图像转换为灰度图像,以降低后续计算量.

针对面部图像中可能存在姿态差异的问题,需要对面部检测结果进行校正. 本文利用旋转变换矩阵对面部姿态差异进行校正,定义如下:

(x',y',1)=[x,y,1]⎡cosθsinθ0−sinθcosθ0001⎤. (1)

式中,x,y 表示原始坐标,x',y'表示变换后的坐标,θ表示旋转角度,该角度通过测量两眼位置向量与水平向量的夹角得到,本研究使用 Cheng 等人提出的 DRMF 框架检测双眼位置^[25]进而计算角度 θ. 如图 2 所示为深度图像面部检测结果及其在灰度图中的映射,本文使用热度图更加清晰地表示深度图像.

(2)降采样
为了减小待处理图片的尺寸,保证输入数据的尺度一致,需要对不同模态图像降采样. 本文采用线性插值方法进行降采样,降采样后尺寸为 128 * 164(单位:像素).

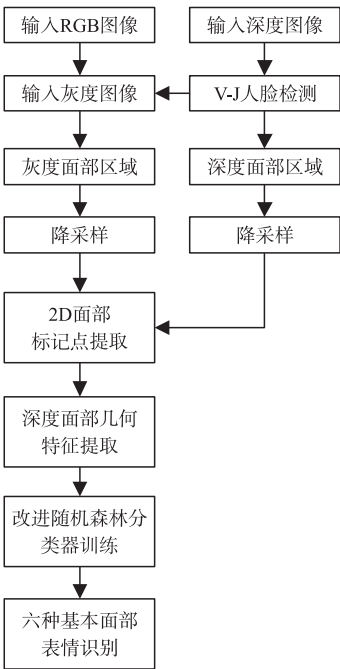


图 1 基于 RGBD 面部图像与改进随机森林分类器的表情识别流程
Fig. 1 The flow diagram of expression recognition based on bimodal facial images and improved random forest classifier

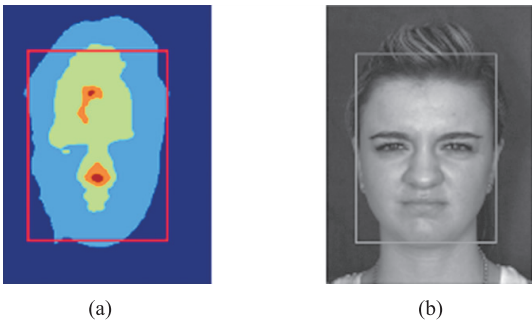


图 2 面部图像检测结果
Fig. 2 Detection results of facial image

2.3 深度面部几何特征

不同于面部检测,面部表情变化通常只涉及局部区域,因此几何特征具有更强的鲁棒性. 本文采用条件局部神经场(conditional local neural fields, CLNF)^[26]分别检测被试特定表情与中性表情的面部标记点并进行跟踪,同时利用不同表情下对应标记点之间的距离差异、角度变化构造 2D 面部几何特征. 本研究借鉴文献[6]中面部运动单元的定义来筛选面部标记点,仅提取与面部表情变化高度相关的点(如脸部、眼睛、鼻子、嘴等部位的轮廓点). 令 2D 面部标记点构成集合 $P_2 = \{p'_i = (x_i, y_i), i = 1, \dots, V\}$, 其中 V 表示标记点数目(本文令 $V = 68$), (x_i, y_i) 表示标记点位置坐标.

2D 面部图像易于获取,但是,缺乏深度信息可能导致 2D 面部几何特征难以准确识别细微表情变化. 3D 面部图像可以提供准确的几何特征,但是 3D 面部图像重构计算量大,较难实时应用. 本文有效利用 RGBD 面部图像的互补性,向 2D 面部几何特征中添加深度信息,以较低成本改善了所提出几何特征识别细微表情变化的能力. 具体做法为:将 2D 面部标记点集合 P_2 中所有点映射到对应的深度图像,构成 3D 面部标记点集合 $P_3 = \{p_i = (x_i, y_i, z_i), i = 1, \dots, V\}$, 其中 z_i 表示点 (x_i, y_i) 的深度(图 3 表示“沮丧”与“惊讶”表情下的面部特征点提取及这些点映射到深度模态图像中的位置,为了显示清楚,面部图像用 RGB 图像表示). 假设被试在表情图像 A 中检测到点集 P_{3A} , 对应的中性表情图像 N 中检测到点集 P_{3N} , 对于 P_{3A} 中任意两点 p_{iA} 以及 p_{jA} , P_{3N} 中有对应的 p_{iN} 以及 p_{jN} , 那么距离差异 Δd_{AN}^{ij} 定义如下:

$$\Delta d_{AN}^{ij} = d_{i,jA} - d_{i,jN} = \sqrt{(x_{iA} - x_{jA})^2 + (y_{iA} - y_{jA})^2 + (z_{iA} - z_{jA})^2} - \sqrt{(x_{iN} - x_{jN})^2 + (y_{iN} - y_{jN})^2 + (z_{iN} - z_{jN})^2}. \quad (2)$$

角度变化 θ_{AN}^{ij} 计算方式如下:

$$\cos \theta_{AN}^{ij} = \frac{p_{iA} p_{jA} \cdot p_{iN} p_{jN}}{|p_{iA} p_{jA}| \cdot |p_{iN} p_{jN}|} = \frac{(x_{iA} - x_{jA}, y_{iA} - y_{jA}, z_{iA} - z_{jA}) \cdot (x_{iN} - x_{jN}, y_{iN} - y_{jN}, z_{iN} - z_{jN})}{|(x_{iA} - x_{jA}, y_{iA} - y_{jA}, z_{iA} - z_{jA})| \cdot |(x_{iN} - x_{jN}, y_{iN} - y_{jN}, z_{iN} - z_{jN})|}. \quad (3)$$

根据上述方法计算点集 P_{3A} 中任意两点的距离差异与角度变化,并拼接所有结果构成表情图像 A 的深度面部几何特征. 由于面部标记点总数为 68, 因此得到的几何特征维度为 $C_{68}^2 \times 2 = 4556$. 也有研究者在 68 个标记点中继续寻找能够有效描述表情变化的子集, 本文使用全部 68 个标记点以避免遗漏面部表情信息. 不难看出,添加深度信息后,面部几何特征的距离差异与角度变化中包含了更多信息,能够更好地描述面部细微表情变化. 在后续实验分析中,将验证深度信息对算法性能的改善.

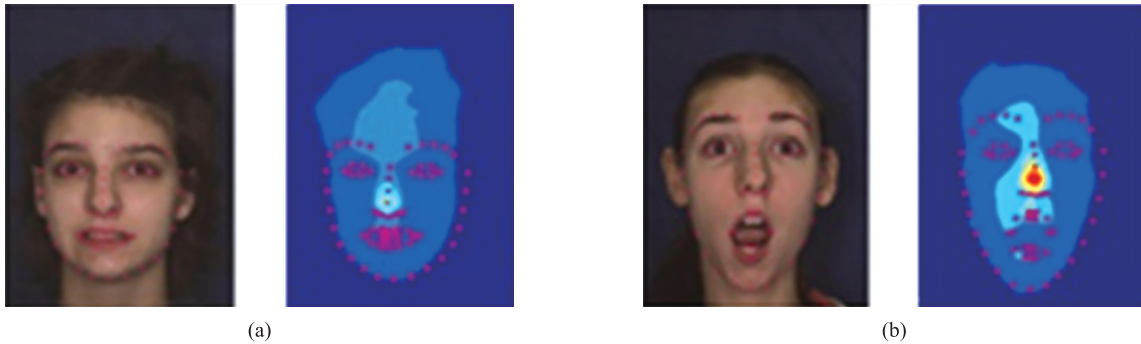


图 3 (a)“沮丧”表情及(b)“惊讶”表情面部标记点在 RGB 图像上的提取及其在深度图像上的映射

Fig. 3 Facial expression of sadness and surprise extract on RGB image and their mapping on depth image

2.4 改进随机森林分类器

本文采用随机森林(RF)分类器对面部表情进行分类. 假设特征 O 的长度为 M (M 表示输入特征维数), RF 分类器构造 K 棵决策树 $\{T_i(O)\} (i = 1, \dots, K)$, 并对所有决策树的分类结果进行投票, 从而输出最终分类结果. RF 分类器的最终分类结果为:

$$f_{RF}(O) = \frac{1}{K} \sum_{i=1}^K T_i(O). \quad (4)$$

RF 分类器通过随机选择样本与特征保证了算法的泛化能力. 但是,并非所有特征都能为分类提供等价的信息,某些特征可能会导致错误的分类. 从 3.3 深度面部几何特征的提取过程中可以看出,最终得到的 4556 维特征向量并非都对面部表情识别有均等贡献,譬如眼角标记点与下巴标记点在某表情下的距离差异与角度变化(相对于中性表情图像中的对应点)对识别该表情贡献不大,因此需要将这对特征剔

除,防止分类器的错误学习. 为了避免无关特征影响 RF 分类器的性能,本文对输入特征 O 进行筛选,设计基于特征选择的 RF 分类器(feature selection based random forest,FSRF). 特征筛选过程如下:对于输入特征 O 中任意元素 $O_j(j=1,\cdots,M)$,其重要性得分 IP_j 计算方式如下:

$$IP_j = \frac{1}{K_j} \sum_{v \in S_j} \text{Gain}(O_j, v). \quad (5)$$

式中, S_j 表示随机森林中 O_j 分割的 K_j 棵决策树的节点集合. $\text{Gain}(O_j, v)$ 表示由 Gini 指数计算得到的信息增益. 本文采用 O 中所有元素重要性得分的均值作为阈值来筛选特征 O ,即将 O 中重要性得分大于阈值的元素构成新的特征 O^* ,并将 O^* 代入式(4)进行 RF 分类器训练.

2.5 算法复杂度分析

本文算法中人脸检测通过经典 Viola-Jones 算法实现,对于检测得到的人脸,提取深度面部几何特征并利用改进的随机森林进行分类,这部分算法的计算复杂度如下:令样本个数为 m ,特征维数为 M ,用于投票的决策树数目为 K ,那么随机森林算法的计算复杂度为 $O(K * M * n * \log(n))$. 对于每个特征,需要通过整合每个元素在点集 S_j 中的信息增益来计算该元素的重要性得分,计算复杂度为 $O(M * n * K_j)$. 另外,提取深度面部几何特征与样本数目 n 成线性关系,因此其复杂度为 $O(n)$. 综上所述,本文提出的特征提取及表情识别算法的计算复杂度为:

$$O(K * M * n * \log(n)) + O(M * n * K_j) + O(n) \approx O(K * M * n * \log(n)). \quad (6)$$

不难看出,计算复杂度与特征维数 M 成线性关系,与样本个数 n 的平方成近似线性关系(K 的大小相对于 M 、 n 可忽略),因此算法在处理高维特征小样本问题时运算时间可接受.

3 实验结果分析

3.1 标准数据库及实验环境

本文在 VS2013+OpenCV3.0 开发环境下对所提算法进行实验分析,硬件配置为 Intel i7 处理器+12G 内存+2G 显存. 本文利用“VT-KFER database”与“CurtinFaces database”2 个面部 RGBD 基准数据库对算法进行测试,同时,通过 Kinect 2.0+VS2013+OpenCV3.0 搭建了一个在线面部表情识别系统,用来评价算法在复杂环境下的有效性.

(1)“VT-KFER database”^[27]:该数据库由 kinect 1.0 采集到的 RGBD 面部表情图片组成,包含时序信息,带标记. 数据库样本总数为 1 956 个 RGB 图像序列,每个序列表示一种表情变化,每个序列平均包含 6 帧图片. 数据采自 32 个被试(男/女,年龄 10 岁~30 岁),包含 6 种基本表情(高兴、伤心、惊讶、沮丧、恐惧、愤怒)以及中性表情. 在实验中,从每种基本表情中各选用 200 个样本进行,一共选取了 1 400 个样本.

(2)“CurtinFaces database”^[28]:该数据库由 kinect 1.0 采集到的 RGBD 面部表情图片组成,带标记. 数据样本总数超过 5 000,采自 52 个被试(男/女,戴眼睛/不戴眼镜),包含 6 种基本表情(高兴、伤心、惊讶、沮丧、恐惧、愤怒)以及中性表情,包含不同视角、不同光照、遮挡等实际干扰. 在实验中,从每种基本表情中各选用 500 个样本进行,一共选取了 3 500 个样本.

(3)“实际数据集”:课题组基于 Kinect 2.0 采集了 28 个被试(男/女,戴眼睛/不戴眼镜)的 6 种基本表情以及中性表情的面部 RGBD 图像,数据样本总数为 2 300. 在采集面部表情时考虑了头部无旋转/头部有旋转以及部分遮挡等干扰因素,暂未考虑光照变化因素. 在 RGBD 数据的实际采集过程中,由于 Kinect 设备在采集 RGB 图像(以 RGB 摄像头为原点)与深度图像(以红外摄像头为原点)时采用了不同的空间坐标系,导致成像存在差异,因此在识别面部表情前,需要采用“Kinect for windows”SDK 中的映射函数 MapDepthToColorImagePoint 完成不同模态图像的对齐.

3.2 定量实验分析

(1) 面部表情识别率分析

图 4(a)、(b)、(c)分别给出了本文算法在“VT-KFER”数据库、“CurtinFaces”数据库以及实际数据集上对 6 种基本面部表情的识别结果. 对于任意数据集,选用 70%作为训练样本,30%作为测试样本,并随机重复 10 次实验,以混淆矩阵形式给出平均结果.

对于“VT-KFER”数据库,算法对 6 种基本表情的平均识别率为 0.88~0.96,均值为 0.922,标准差为

angry	0.95	0.92	0.01	0.90	0.02	0.00
disgust	0.03	0.88	0.03	0.00	0.04	0.02
fear	0.01	0.03	0.92	0.00	0.01	0.03
happy	0.02	0.00	0.00	0.96	0.00	0.02
sadness	0.00	0.05	0.03	0.02	0.88	0.02
surprise	0.03	0.00	0.02	0.01	0.00	0.94

(a) “VT-KFER” 数据库

angry	0.93	0.02	0.02	0.00	0.01	0.02
disgust	0.01	0.89	0.05	0.00	0.04	0.01
fear	0.02	0.03	0.90	0.01	0.01	0.03
happy	0.02	0.00	0.04	0.92	0.00	0.02
sadness	0.01	0.08	0.02	0.02	0.85	0.02
surprise	0.02	0.00	0.02	0.01	0.01	0.94

(b) “CurtinFaces” 数据库

angry	0.85	0.05	0.04	0.03	0.01	0.02
disgust	0.03	0.78	0.10	0.02	0.04	0.03
fear	0.03	0.08	0.81	0.02	0.02	0.04
happy	0.03	0.01	0.06	0.88	0.00	0.02
sadness	0.01	0.08	0.02	0.02	0.86	0.01
surprise	0.04	0.00	0.03	0.02	0.00	0.91

(c) 实际数据集

图 4 基本表情识别率 (a) “VT-KFER” 数据库, (b) “CurtinFaces” 数据库和 (c) 实际数据集

Fig. 4 Recognition rate of basic expression (a) “VT-KFER” database, (b) “CurtinFace” database and (c) real database

0.035; 对于“CurtinFaces”数据库, 平均识别率为 0.85~0.94, 均值为 0.905, 标准差为 0.033; 对于实际数据集, 平均识别率为 0.78~0.91, 均值为 0.848, 标准差为 0.047. 从结果不难看出, 由于从“VT-KFER”数据库中挑选出的面部表情样本中不包含影响表情识别的因素 (如戴眼镜、面部遮挡等), 因此识别准确率较高; 从“CurtinFaces”数据库中挑选出的面部表情样本包含不同视角、不同光照、遮挡等实际干扰, 因此识别准确率略低; 而实际数据集中包含一些严重的人为干扰 (大面积脸部遮挡或大角度头部旋转), 因此识别率最低.

针对所有数据库的表情识别结果, 可以得出如下结论: 算法对“沮丧”与“悲伤”两种表情的识别率较低, 而对“高兴”与“惊讶”两种表情的识别率较高. 这是由于前两种表情牵涉的面部运动较小, 因此相应的距离差异与角度变化较小, 导致深度面部几何特征对表情识别率偏低, 而后两种表情牵涉到较大的面部运动, 因此相应的距离差异与角度变化较大. 总的来说, 本文算法可以在复杂环境下较为准确地识别 6 种基本表情.

(2) 算法有效性分析

算法有效性主要探讨以下两个问题: ①引入深度信息构造的深度面部几何特征是否优于 2D 面部几何特征; ② RF 分类器中的特征选择策略是否有效. 为了回答上述两个问题, 本文分别在“VT-KFER”与“CurtinFaces”基准数据库上进行测试. 对于任一数据库, 将所有样本的 70% 作为训练样本, 30% 作为测试样本, 并随机重复 10 次实验, 表 1、表 2 分别给出了两组实验结果, 用表情识别准确率的平均值与标准差进行衡量.

从表 1 结果不难看出, 通过添加额外的深度信息, 深度面部几何特征对不同表情的识别率明显优于 2D 面部几何特征, 深度方法的平均识别准确率更高, 而且标准差更小, 因此算法识别不同表情的鲁棒性更强.

表 2 结论说明了引入特征选择策略对 RF 分类器的影响, 通过对输入的大量特征进行筛选, 可以有效排除与表情识别无关的特征, 避免这些特征影响分类器的性能, 从而提高分类器的平均识别准确率, 减小标准差.

(3) 不同算法效果对比分析

为了进行更全面的分析评价, 本文在“VT-KFER”与“CurtinFaces”基准数据库上与主流面部表情识别算法进行了对比. 同样将所有样本的 70% 作为训练样本, 30% 作为测试样本, 随机重复 10 次实验. 对比算法中的参数依照原文献进行设置. 表 3 给出了不同算法对 6 种基本表情的平均识别准确率, 不难看出本文算法表现优于基于手动提取特征的表情识别算法^[29-31], 性能接近基于卷积神经网络的表情识别算法^[32]. 但是, 本文算法在训练阶段对样本数量以及硬件设备的要求远低于基于卷积神经网络的方法, 因此更容易推广.

表 1 深度/2D 面部几何特征表情识别结果比较

Table 1 Compare depth with 2D facial geometry feature expression recognition results

面部表情	VT-KFER		CurtinFaces	
	深度面部几何特征	2D 面部几何特征	深度面部几何特征	2D 面部几何特征
angry	95±1.3	91±2.3	93±1.8	89±2.6
disgust	88±3.2	78±4.1	89±2.2	75±3.8
fear	92±2.1	81±3.3	90±2.1	82±3.1
happy	96±1.1	94±1.9	92±1.7	90±1.9
sadness	88±2.8	82±3.8	85±2.7	84±2.8
surprise	94±1.5	92±2.8	94±1.5	92±2.3

表 2 FSRF/RF 分类准确率对比					表 3 主流面部表情识别算法平均准确率对比		
Table 2 FSRF/RF classification accuracy comparison					Table 3 Famous facial expression recognition algorithm average accuracy rate comparison		
面部表情	VT-KFER		CurtinFaces		采用算法	数据库	
	FSRF	RF	FSRF	RF		VT-KFER	CurtinFaces
angry	95±1.3	94±2.1	93±1.8	91±2.3	Fan et al.[30]	84.70%	82.11%
disgust	88±3.2	83±4.5	89±2.2	84±3.9	Rivera et al.[31]	91.51%	88.75%
fear	92±2.1	88±3.5	90±2.1	86±3.4	Aly et al.[29]	88.14%	87.32%
happy	96±1.1	94±2.3	92±1.7	92±2.1	Zhang et al.[32]	95.12%	93.48%
sadness	88±2.8	85±3.8	85±2.7	85±3.0	本文方法	92.23%	90.52%
surprise	94±1.5	94±1.8	94±1.5	90±2.5			

3.3 实际图片检测结果

图 5 所示为本文算法在实际数据集上的识别结果,其中红色矩形框表示面部检测结果,矩形框上方为表情识别结果及预测概率. 图 5(a) 中的表情为“surprise”,算法以 0.94 的概率成功识别了该表情;图 5(b)、(c)检测到的面部中包含部分遮挡,由于面部几何特征对遮挡有一定鲁棒性,算法仍能够以 0.85 与 0.84 的概率成功识别这两个表情(“surprise”与“happy”);图 5(d)中的表情为“angry”,但是算法以 0.57 的概率将该表情错分为“sadness”,导致这种错误的原因是该图像中被试的表情比较微弱.



图 5 实际数据集面部表情识别结果
Fig. 5 Recognition results of real dataset

4 结束语

本文提出了一种基于 RGBD 面部图像与改进随机森林分类器的表情识别算法. 算法利用深度面部图像对光照变化鲁棒的特性,在其进行面部检测;根据面部检测结果,在灰度图上检测并跟踪与表情变化高度相关的面部标记点,并从对应的深度图像上提取深度信息,构造深度面部几何特征,相比于 2D 面部图像几何特征,前者能够更好地识别面部表情变化,同时不需要进行复杂的 3D 面部图像重构;考虑到该特征中并非所有元素都对分类器的训练有均等贡献,引入了特征选择策略改善随机森林分类器性能. 本文在“VT-KFER”与“CurtinFaces”基准数据库上进行实验并与几种主流表情识别算法进行对比,结果验证了本文算法的有效性,其识别准确率高于传统基于手动提取特征的表情识别算法,性能接近基于深度学习的表情识别方法. 但是,本文算法对训练样本数量及硬件平台要求远低于深度学习方法,因此更容易推广. 本文搭建了基于 Kinect 2.0 的面部表情识别系统,并在复杂情况下验证了算法性能,后续研究方向主要是如何对微弱面部表情变化进行准确识别.

[参考文献]

[1] 黄建,李文书,高玉娟. 人脸表情识别研究进展[J]. 计算机科学,2016,43(S2):123-126.
[2] WU S Z,KAN M N,HE Z L,et al. Funnel-structured cascade for multi-view face detection with alignment-awareness[J]. Neurocomputing,2017,221:138-145.
[3] DEMIRKUS M,PRECUP D,CLARK J,et al. Multi-layer temporal graphical model for head pose estimation in real-world videos[C]//International Conference on Image Processing. China:IEEE,2014:3392-3396.
[4] JAIN S,HU C B,AGGARWAL J K. Facial expression recognition with temporal modeling of shapes[C]//The 13th International Conference on Computer Vision. Spain:IEEE,2011:1642-1649.
[5] ALI G,IQBAL M A,CHOI T S. Boosted NNE collections for multicultural facial expression recognition[J]. Pattern recognition,2016,55:14-27.
[6] ZHANG Y,ZHANG L,HOSSAIN M A. Adaptive 3D facial action intensity estimation and emotion recognition[J]. Expert systems with applications,2015,42(3):1446-1464.

- [7] MOEINI A,FAEZ K,SADEGHI H,et al. 2D facial expression recognition via 3D reconstruction and feature fusion[J]. Journal of visual communication and image representation,2016,35:1-14.
- [8] MOHAMMADI M,FATEMIZADEH E,MAHOOR M. Pca-based dictionary building for accurate facial expression recognition via sparse representation[J]. Journal of visual communication and image representation,2014,25(5):1082-1092.
- [9] 首照宇,杨晓帆. 联合 Gabor 误差字典和低秩表示的人脸识别算法[J]. 计算机科学,2017,44(3):296-299.
- [10] SHAN C,GONG S,MCOWAN P W. Facial expression recognition based on local binary patterns:a comprehensive study[J]. Image and vision computing,2009,27(6):803-816.
- [11] 蒋政,程春玲. 基于 Haar 特性的改进 HOG 的人脸特征提取算法[J]. 计算机科学,2017,44(1):303-307.
- [12] KOBAYASHI H,HARA F. Facial interaction between animated 3d face robot and human beings[C]//IEEE International Conference on Computational Cybernetics and Simulation. USA:IEEE,1997:3732-3737.
- [13] KENJI M. Recognition of facial expression from optical flow[J]. Ieice transactions on information and systems,1991,74(10):3474-3483.
- [14] VALSTAR M,PANTIC M,PATRAS I. Motion history for facial action detection in video[C]//IEEE International Conference on Systems,Man and Cybernetics. Netherland:IEEE,2004:635-640.
- [15] ZHAO G,PIETIKAINEN M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. IEEE transactions on pattern analysis and machine intelligence,2007,29(6):915-928.
- [16] ZHAO X M,SHI X G,ZHANG S Q. Facial expression recognition via deep learning[J]. IETE technical review,2015,32(5):347-355.
- [17] 卢官明,何嘉利,闫静杰,等. 一种用于人脸表情识别的卷积神经网络[J]. 南京邮电大学学报(自然科学版),2016,36(1):16-22.
- [18] YU Z D,ZHANG C. Image based static facial expression recognition with multiple deep network learning[C]//Proceedings of the 2015 Acm International Conference on Multimodal Interaction. Netherland:ACM,2015:435-442.
- [19] FAN X,JIA Q,HU Y K,et al. 3D facial landmark localization using texture regression via conformal mapping[J]. Pattern recognition letters,2016,83:395-402.
- [20] DANELAKIS A,THEOHARIS T,PRATIKAKIS I. A spatio-temporal wavelet-based descriptor for dynamic 3D facial expression retrieval and recognition[J]. The visual computer,2016,32(6):1001-1011.
- [21] KUNG S H,ZOHDY M A,BOUCHAFFRA D. 3D HMM-based facial expression recognition using histogram of oriented optical flow[J]. Transactions on machine learning and artificial intelligence,2016,3(6):42.
- [22] ZHEN Q,HUANG D,WANG Y,et al. Muscular movement model-based automatic 3D/4D facial expression recognition[J]. IEEE transactions on multimedia,2016,18(7):1438-1450.
- [23] ZHANG Z,CUI L,LIU X,et al. Emotion detection using Kinect 3D facial points[C]//International Conference on Web Intelligence. USA:IEEE,2016:407-410.
- [24] BARNOUTI N H,ALDABBAGH S S M,MATTI W E,et al. Face detection and recognition using Viola-Jones with PCA-LDA and square euclidean distance[J]. International journal of advanced computer science and applications,2016,7(5):371-377.
- [25] CHENG S,ASTHANA A,ZAFEIRIOU S,et al. Real-time generic face tracking in the wild with cuda[C]//Proceedings of the 5th ACM Multimedia Systems Conference. USA:ACM,2014:148-151.
- [26] BALTRUSAITIS T,ROBINSON P,MORENCY L P. Constrained local neural fields for robust facial landmark detection in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Australia:IEEE,2013:354-361.
- [27] LI B,MIAN A,LIU W,et al. Using kinect for face recognition under varying poses,expressions,illumination and disguise[C]//IEEE Workshop on Applications of Computer Vision. Florida:IEEE,2013:186-192.
- [28] ALY S,TRUBANOVA A,ABBOTT L,et al. VT-KFER:A Kinect-based RGBD+Time dataset for spontaneous and non-spontaneous facial expression recognition[C]//International Conference on Biometrics. USA:IEEE,2015:90-97.
- [29] ALY S,ABBOTT A L,TORKI M. A multi-modal feature fusion framework for kinect-based facial expression recognition using Dual Kernel Discriminant Analysis(DKDA)[C]//IEEE Winter Conference on Applications of Computer Vision. USA:IEEE,2016:1-10.
- [30] FAN X,TJAHJADI T. A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences[J]. Pattern recognition,2015,48(11):3407-3416.
- [31] RIVERA A R,CASTILLO J R,CHAE O O. Local directional number pattern for face analysis:Face and expression recognition[J]. IEEE transactions on image processing,2013,22(5):1740-1752.
- [32] ZHANG W,ZHANG Y M,MA L,et al. Multimodal learning for facial expression recognition[J]. Pattern recognition,2015,48(10):3191-3202.