

文章编号: 2095-2163(2022)11-0258-04

中图分类号: TP319

文献标志码: A

基于表情识别技术的电影效果反馈

于初允, 李 闯

(吉林师范大学 计算机学院, 吉林 四平 136000)

摘要: 表情识别是人脸识别的重要分支,具有客观性和敏捷性的特点。市场上关于电影的评价,多是由人的主观意志来给出评分,缺乏客观性,电影效果的反馈缺少数据支撑。基于表情识别技术的电影效果反馈,可以为电影效果反馈提供客观数据,保证其可靠性,实现了情绪向数据的转变。利用 Dlib 的特征提取器捕获人脸,利用 Dlib 的特征预测器提取人脸的 68 特征点进行几何计算^[1],达到表情识别的效果。通过实验,检测到实验者在观看电影时的自然表情、高兴、惊讶、生气和恐惧五种表情。本系统可以迅速、客观并且较为准确地分析电影观看者的表情,进行反馈数据的统计。

关键词: 人脸识别; 表情识别; 电影效果反馈

Movie effect feedback based on expression recognition technology

YU Chuyun, LI Chuang

(College of Computer Science, Jilin Normal University, Siping Jilin 136000, China)

【Abstract】 Expression recognition is an important branch of face recognition, and has the characteristics of objectivity and agility. The scoring of films in the market is mainly carried out by people's subjective intension, which is lack of objectivity, meanwhile the feedback of film effects is lack of data support. Film effect feedback based on expression recognition technology can provide objective data for film effect feedback, ensure its reliability, and realize the transformation from emotion to data. The feature extractor of Dlib is used to capture the face, and the feature predictor of Dlib is used to extract 68 feature points of the face for geometric calculation^[1], so as to achieve the effect of expression recognition. Through the experiment, five expressions such as natural expression, happiness, surprise, anger and fear are detected when the experimenter watches the film. The system can quickly, objectively and accurately analyze the expression of film viewers and make statistics of feedback data.

【Key words】 face recognition; expression recognition; film effect feedback

0 引言

伴随科学技术迅速发展,门禁、手机解锁和快捷支付等人脸识别技术的应用,给人们生活带来极大便利,人脸识别逐渐走进了日常生活中。而经济发展也丰富了人们的精神文化生活,如看电影就已经成为了人们休闲娱乐的重要选择之一。但互联网上关于影评褒贬不一,都是凭借人的主观意志对影片进行评价,人们获取有效信息的效率降低,挑选合适电影付出的时间代价大。表情识别是人脸识别的重要组成部分,技术日趋成熟且效率较高、跨平台性能优异。通过对观影者在观看电影过程中的表情识别,与预期效果做出对比,达到对电影效果进行客观反馈的目的。通过多种实验验证,基于 Dlib 的表情识别技术可以对电影进行有效反馈。

1 Dlib 技术介绍

1.1 ResNet

ResNet(Residual Neural Network),即基于残差学习思想的深度神经网络,将人脸数据的训练集与面部特征数据进行对比,是 Dlib 在人脸识别部分的核心^[2]。

残差学习结构如图 1 所示。将学习到的特征记为 $H(x)$,由于残差学习相较于原始特征的学习更加容易,由其学习到的残差为 $F(x) = H(x) - x$,通过不断地进行 $F(x) + x$ 的学习,使系统具有更佳性能^[3]。

ResNet 有 2 种结构:2 层残差学习单元和 3 层残差学习单元,如图 2 所示。2 种结构分别对应 ResNet34 和 ResNet50/101/152,后者相较于前者显著的优点是减少了参数量。所以前者常用于层数更少的网络,后者常用于更深的网络^[3]。

基金项目: 第十九批大学生科研基金项目(19043)。

作者简介: 于初允(2001-),女,本科生,主要研究方向:人像识别;李 闯(1982-),女,博士,副教授,主要研究方向:人工智能、数据库技术应用。

通讯作者: 李 闯 Email:14659302@qq.com

收稿日期: 2022-03-20

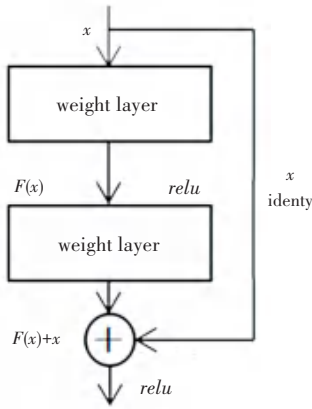


图 1 ResNet 的残差学习结构^[3]

Fig. 1 Residual learning structure of ResNet^[3]

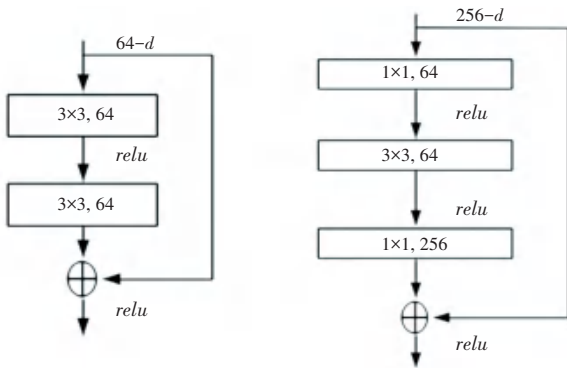


图 2 ResNet 的 2 种结构^[3]

Fig. 2 Two structures of ResNet^[3]

Dlib 使用的是 34 层的网络, ResNet34 的最后一层是 fc 1000, 为得到 128 维向量, 在 fc 1000 后面加 Dense(128)。通过获取人脸的 Dense(128) 与原 Dense(128) 的欧式距离, 与阈值和相似程度的百分比做比较, 即可判断获得到的人脸图像信息。

1.2 Dlib68 特征点

本文采用 Dlib 的 68 特征点检测模型对人脸进行检测, 特征点位置如图 3 所示。



图 3 Dlib 的 68 特征点

Fig. 3 68 feature points of Dlib

此模型的训练流程为: 首先, 将训练图片和测试

图片放于同一文件目录下, 为保证此模型识别准确率的提升, 应使训练集和测试集图片包含多种可能, 如不同光照、距离的图片, 为扩大两集合, 做镜像处理。其次, 图像中的人脸尺寸可能存在过大差异, 影响结果输出, 所以应设置适当扫描窗口。使用 SVM 训练器训练所提取特征^[4], 可得到人脸检测模型。最后, 测试人脸检测模型, 若达到预期标准, 即可应用。训练过程如图 4 所示。

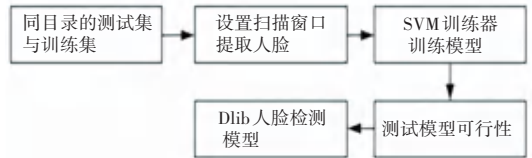


图 4 Dlib68 特征点的训练过程

Fig. 4 Training process for Dlib68 feature points

2 具体实现

2.1 系统框架

本文应用的表情识别框架如图 5 所示, 模块应用了人脸图像的特征点检测和简单的表情计算算法, 简单概括为: 首先, 为获取人脸图像使用人脸检测技术, 判断是否为人类面部; 其次, 为减少环境带来的影响, 对检测到的人脸图像进行预处理; 再次, 对于面部的特征检测, 需要用到人脸图像的特征点检测技术; 最后, 通过特征点计算进行简单的表情识别。将表情识别的结果与目标电影片段的预期效果进行对比, 实现电影效果反馈。

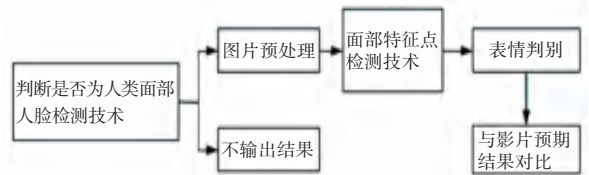


图 5 系统识别框架

Fig. 5 System identification framework

2.2 实现步骤

论文中重点研究了通过利用 Dlib 特征提取器和特征预测器对人脸进行捕获和特征提取, 以达到简单的表情识别。设计实现步骤具体如下:

(1) 利用 OpenCv 捕获图像^[5], 并对每帧图像进行灰度处理。

(2) 利用 Dlib 的特征提取器检测人脸, 并计算数量。

(3) 若存在检测到的人脸数不为空, 显示每个人脸的 68 特征点, 并计算人脸识别框的长度。使用特征预测器获得 68 点数据的坐标。

(4)通过步骤(3)获得有关嘴、眉毛和眼睛的坐标数据,进行简单的表情计算。需要获得的坐标有:嘴中心、嘴左角和嘴右角;眉毛的10个特征点;2只

眼睛的眼角、眼尾、上眼睑中点及下眼睑中点。

(5)通过步骤(4)获得的坐标进行计算,判断表情。表情判别流程如图6所示。

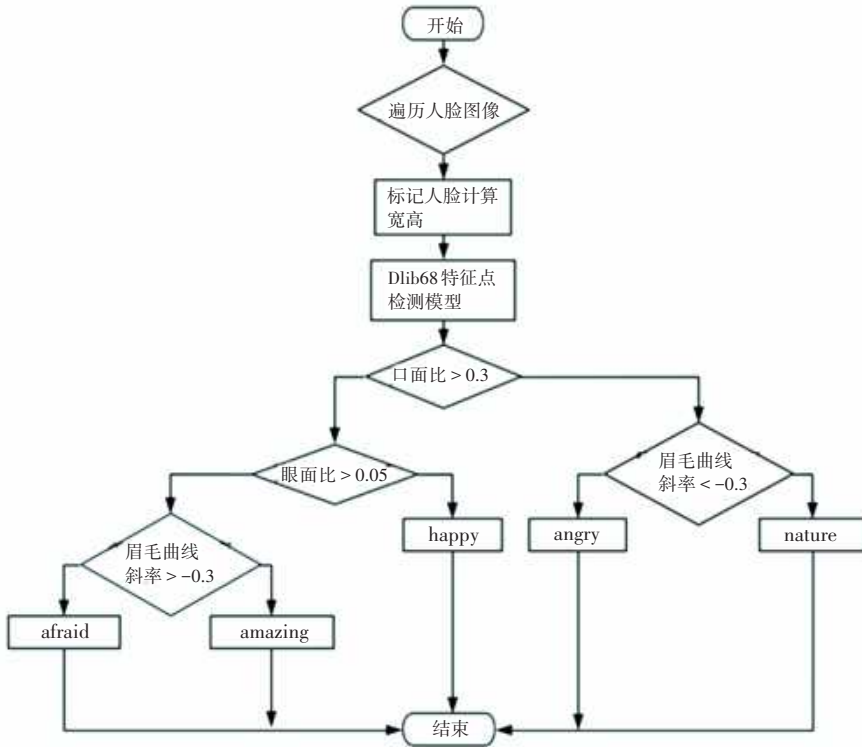


图6 表情判别流程图

Fig. 6 Expression discrimination flowchart

2.3 存在问题

综合前文提及的表情识别中基于面部几何的识别方法^[7]可知,利用检测获得的68特征点,通过数学计算,对人在恐惧、惊讶、高兴、愤怒和自然状态下进行表情的识别。进而可以通过获得的表情数据,与预期观众的观影效果进行对比,达到电影效果反馈的目的。研究发现在此过程中存在的主要问题有:

(1)电影中呈现的内容复杂繁多,不同的人所感受到的情绪不同,所展现的表情各异。期望情绪集的数量偏大,会导致不符合预期的表情也被计入预期结果中,导致反馈不准确;数量偏小,则符合预期的表情会有大部分不计入预期结果中,导致反馈不准确。所以期望情绪集的创建不具备便捷、准确的优点。

(2)由于人的个体个性、身体发育等方面存在差异,对于同种情绪的表现也有所不同,表现程度也并不一样。且存在调动的肌肉、组织变化细微,多数呈现复合表情,约超过21种。本文探讨的表情识别,需要在表情变化幅度较大时才能准确识别,缺少微表情的识别。且仅对嘴、眼睛和眉毛的部分变化

作为表情识别的依据,缺少准确性且识别模式单一,并且仅能识别5种普通表情。

2.4 实验步骤及结果分析

为验证上述系统是否能对观影人的表情及时、准确地进行识别,对正在观影的人的表情进行简单识别。测试此系统的可行性,并将结果输出在屏幕中。对该过程可做阐释分述如下。

- (1)选择3类短片:喜剧、恐怖和悬疑。
- (2)开启系统,调用pc摄像头捕获测试者面部。
- (3)记录输出在屏幕上的检测结果。
- (4)测试非人类的面部。
- (5)记录结果。实验结果如图7所示。

研究可知,在喜剧时可以检测到实验者有较频繁的“happy”表情,悬疑时有“amazing”表情和“angry”表情,恐怖时可以检测到实验者的“afraid”表情和部分“angry”表情,大多时候人们都是“nature”表情。基于Dlib的表情识别技术可以较为准确和迅速地识别观影者的表情变化,令数据更加客观,说明将其应用于电影效果反馈上是可行的。