文章编号：2095-2163(2019)03-0063-06 中图分类号：TP311.5 文献标志码：A

基于表情分析和视线追踪的用户反馈采集技术

王宁致[1]，黄碧玲[2]，郑敏仪[3]
（1 华南师范大学 国际商学院，广东 佛山 528000；2 华南师范大学 城市文化学院，广东 佛山 528000；
3 广东工业大学 计算机学院，广州 510006）

摘 要：用户反馈，是指使用某一产品的用户对其产品所提出的有关于产品的情况反馈。用户反馈采集有利于公司优化其产品，为用户提供更好的服务。传统的用户反馈采集方法如跨站跟踪、Cookie 跟踪或观察流量信息，仅反馈用户浏览行为的信息，而忽略了用户的潜在兴趣。基于表情分析和眼球视线追踪技术的用户反馈采集核心技术能够反映用户在网页页面浏览时，无意识状态下自然流露的潜在兴趣。研究采用基于类 Haar 特征的面部检测的 Adaboost 算法，及基于深度学习的面部情感识别技术，使人类面部情感识别的正确率可达 90%。同时使用深度学习方法，在没有高精度且昂贵的仪器条件下，仅借助笔记本电脑前置摄像头实现视线追踪的效果。测试比较 3 种不同的深度学习的网络结构实现视线追踪的准确率，其中效果最佳的一种网络结构的准确率可达 49.60%。

关键词：用户反馈采集；视线追踪；表情分析；深度学习

Research on user feedback acquisition based on expression recognition and sight tracking technology

WANG Ningzhi[1], HUANG Biling[2], ZHENG Minyi[3]
(1 International Business College, South China Normal University, Foshan Guangdong 528000, China;
2 School of Urban Culture, South China Normal University, Foshan Guangdong 528000, China;
3 School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

【Abstract】 Customer feedback refers to the feedback of customers who use a product and then give back a comment about the product. Feedback analytics services use customer generated feedback data to measure customer experience and improve customer satisfaction. Feedback data is collected, then, key performance indicators and feedback metrics is turned into actionable information for website improvement. This paper studies the core technology of feedback analytics services based on facial expression recognition and eye tracking for Web page browsing. The research has built a website that has a function to recognize human expression. The accuracy of expression recognition can reach 90%. The research also explores three different deep learning network structures and tests their accuracy of eye tracking. Instead of the help of high precision and expensive instruments, the research obtains the face pictures through the front camera of the laptop computer and uses the depth learning method to analyze the eye tracking. The accuracy of the best network structure among the three can reach 49.60%.

【Key words】 feedback analytics services; eye tracking; facial expression recognition; in-depth learning

0 引 言

企业对用户评价的引导、跟踪与采集，成为用户反馈采集的重要来源。本文研究的是，如何利用深度学习技术在网页页面浏览时实现高效的用户反馈采集。目前，传统的用户反馈采集方法有 2 种，即：利用跨站跟踪和 Cookie 跟踪等手段采集用户网络行为数据；通过分析网站页面流量和各分界面流量来进行页面整改。前者只反馈用户关注的特定方面的信息，而没有帮助用户发现潜在的感兴趣内容；而后者效率低，反馈整改流程时间过长。

因此，本文提出了基于情感识别和视线追踪技术的用户反馈采集核心技术。通过收集表情数据结合定位视线落点，来判断用户对屏幕上某块区域的内容的感兴趣程度，作为用户反馈数据。这种反馈数据不仅能反映用户理性关注的焦点，还能帮助用户发现潜在感兴趣的内容，而且反馈整改流程时间也较快，甚至可以做到实时反馈。

1 表情分析技术

面部情感识别主要有 3 个环节，分别是：面部检测、情感特征提取和情感分类。为了进行面部情感分析，先要抓取前置快照，并预处理图像中的面部数据，包括定位、矫正尺寸等工作。而后从矫正好的面

基金项目：广东大学生科技创新培育专项资金项目(Pdghb0133)。

作者简介：王宁致(1998-)，女，本科生，主要研究方向：财务管理。

收稿日期：2019-03-04

部图像中提取情感特征,提取特征的质量直接关系到下一步分辨的准确程度。最后就是面部情感分类。根据表情特征性质对所属情感类别进行划定。本文采用由美国心理学家 Friesen 和 Ekman 定义的6种基本情感分类:高兴、惊喜、悲伤、厌恶、生气和恐惧(1970)。

1.1 表情分析技术的研究和实现

表情分析是计算机将提取到的面部特征数据输入分类器,完成分类识别,使计算机能够判定下一步程序的过程。主要分为2个部分。首先是机器学习,提取面部图像的 Haar 特征,用 Adaboost 算法,即采用一种基于级联分类模型的分类器来训练模型。这部分研究旨在获取仅含面部的图像。其次,是深度学习,把前述部分获取的表情输入深度学习网络,从而判断情感类别。这个深度学习网络是采用 Cohn-Kanade 数据库作为训练集进行训练的,6种情感平均识别率可达90%。

1.2 面部检测技术的研究

基于 Haar 特征的 Adaboost 算法由于其速率远高于基于像素识别的算法,可以达到实时识别情感的目的。检测面部后,将获取的面部特征输入深度学习网络。这个深度学习网络是参考了《基于深度学习的情感识别方法研究》。研究得到该网络结构如图1所示。

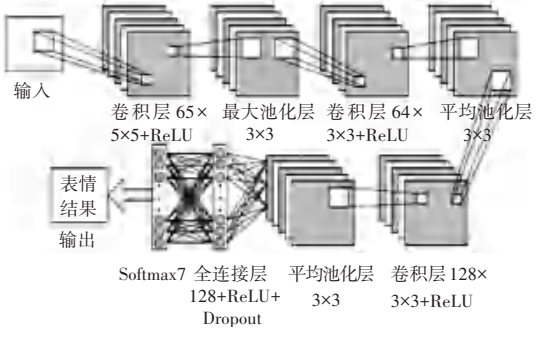

图1 情感识别深度学习网络结构图

Fig. 1 Emotion recognition in-depth learning network structure diagram

本文采用 Cohn-Kanade 数据库训练这个网络。该数据库于2010年发布,其数据量较大、且数据质量较高。这个数据库一共有593张面部图像,本文用总量的75%作为训练集进行训练,即高兴、生气、惊喜、恐惧、厌恶和悲伤六种情感各74张,合计444张图像。用其余149张图像作为测试集进行测试。测试的结果见表1。

表1给出了每种情感的正确识别率,其中对角线的数据就是6种情感分别对应的正确识别率。从实验结果可知,生气、惊喜和悲伤三种情感的识别结果较为理想,其正确识别率都超过了92%,而高兴、恐惧和厌恶三种情感的正确识别率略低,分别为86.10%、87.50%和88.76%。主要原因是高兴与惊喜、恐惧与厌恶相互之间容易产生混淆。

表1 深度学习网络测试结果

Tab. 1 In-depth learning network test results %

	高兴	生气	惊喜	恐惧	厌恶	悲伤
高兴	86.10	0.00	0.00	0.00	0.00	0.00
生气	0.00	92.86	2.44	0.00	0.00	6.71
惊喜	5.63	0.00	94.63	1.32	0.00	0.00
恐惧	4.90	0.00	3.23	87.50	7.79	0.00
厌恶	0.00	3.65	0.00	8.89	88.76	0.00
悲伤	3.37	3.49	0.00	2.29	3.45	93.29

2 视线追踪技术

眼睛是人类获取外界刺激信息的重要视觉器官,大脑中约有80%的知识和记忆都是通过眼睛获取。视线反馈了人类感兴趣的对象、目标和需求,具备输入输出双向性特点。在用户浏览页面时追踪用户视线能获取人机交互的信息,可以采集更多即时的用户反馈,有利于改善传统用户反馈采集的滞后性等缺点。

视线追踪技术是指利用特殊的外接扫描设备获取视线聚焦点的位置和眼球相对于头部位置的运动,并分析注视时间、注视次数、注视顺序和眼眺距离等相关数据,通过终端设备进行处理、计算、分析,最终构建出一个注视点的参考平面图。其技术原理是通过图像传感器采集的角膜反射模式和其他信息,计算出眼球的位置和注视方向。

基于视线追踪技术,市面上已有视线追踪器,或称眼动仪。但眼动仪存在着以下不足:

(1)价格高昂。以瑞典 Tobii 公司为例,最基础的一款 Tobii X2 眼动仪报价硬件和软件共31万人民币。

(2)使用不便。以瑞典 Tobii 公司为例,若要追踪用户视线,用户需额外购买专门的硬件设备和软件程序,使用过程还需要佩戴专门的眼镜。

眼动仪等高成本、需要定制或侵入式硬件以及现实世界中的不准确性等这些因素使得眼动追踪无法成为普通技术。在本文设计中,则拟将使用合理的相机,如智能手机的前置镜头或平板电脑的摄像头,研究一种更物美价廉的视线追踪技术,推广至民用商用,便民利民。

判断情感分析情绪后仍未能得到有效的用户反馈信息。所以结合视线追踪技术继续收集用户的反馈信息，并综合分析这 2 种信息。本文中，研发构建 3 种不同的深度学习的网络结构训练数据集，测试比较并寻求效果最好的网络结构。

2.1　实验数据集

在本次研究中，使用的是来自论文 TabletGaze：dataset and analysis for unconstrained appearance – based gaze estimation in mobile tablets 中公开的数据集。考证该论文后可知，论文中使用了长 22.62 cm、宽 14.14 cm 的屏幕。共有 35 个注视点在平板屏幕上均匀分布，排列成 5 行、7 列，垂直间隔 3.42 cm，左右间隔 3.41 cm。平板屏幕上的注视模式的示例图像即如图 2 所示。

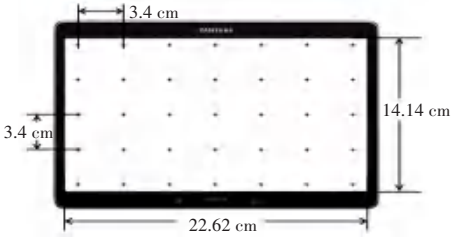

图 2　受试者使用的平板屏幕的示例图像

Fig. 2　Flat-screen sample used by the subjects

此时如图 3 所示，即将显示区域划分成 35 个区域。原始数据是通过平板电脑在景观模式的前置摄像头拍摄受试者的视频得到的，视频采用1 280×720 像素的图像分辨率。受试者在平板屏幕上观察点出现的位置。有 35 个点（5×7）分布在平板电脑屏幕上。在每一个视频里，一个点一次只出现在在一个位置，点的位置在 35 个点之间是随机的，然后随机移动到下一个地方，直到点在所有 35 个位置各出现一次，结束一个视频录制。具体的观察点将遵照图 3 中的数字顺序从小到大依次显示。受试者事先并不知道观察点的显示顺序。

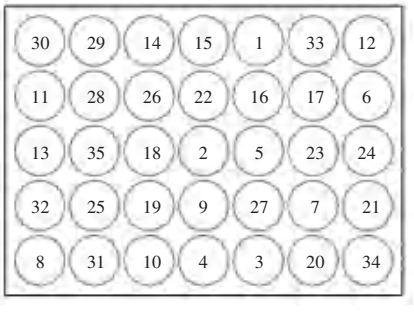

图 3　观测点的显示顺序

Fig. 3　Display order of observation points

该数据集中共有 51 名受试者，12 名女性和 39 名男性参与了数据收集，其中 26 人戴着眼镜；28 名受试者是白种人，其余 23 人是亚洲人。受试者的年龄大约在 20～40 岁之间。每轮数据采集期间，受试者分别用 4 种不同的身体姿势（站、坐，葛优瘫或躺，如图 4 所示）之一，录制一个视频序列。每个主题需要为 4 个身体姿势中的每一个进行 4 次记录，因此每个主题总共收集了 16 个视频序列。

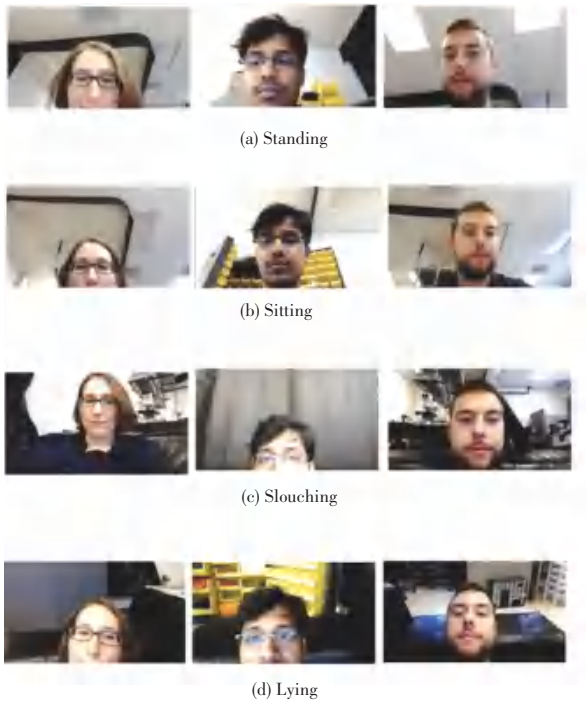

图 4　4 种身体姿势

Fig. 4　Four types of the body

为了使得实验结果有对比性、参考性和可重复性，采用公开的实验数据集进行视线追踪技术的研究。每个视频中每 12 帧截一次图，共取得 143 360 张带面部的截图。将每人每种身体姿势的 70% 的截图，共 100 352 张截图作为训练集训练本文采用的深度学习网络，其余 43 008 张作为测试集检测深度学习网络。

2.2　视线落点定位深度学习网络结构一

本文使用深度学习定位视觉落点。研究中使用的第一种深度学习的网络结构如图 5 所示。

图 5 中各变量的解释见表 2。在此基础上，对该种设计可做研究阐释如下。

（1）输入初始图像。注入初始图像的训练结果如图 6 所示。其中，蓝色线描述的准确率最后只稳定在 33.65%，红色线描述的训练集在模型中的预测结果与真实结果的误差较大。测试结果仅能得知预测的视线落定是否精准定位在测试区域，但无法得到通过深度学习预测的视线落点距离测试区域有多

远。于是通过计算所有 49 115 个预测点和原点的距离差及其平均值,即 2.015 个单位。鉴于本文所采用的屏幕仅有 5×7 个单位,相差 2.015 个单位的测试结果较不理想。

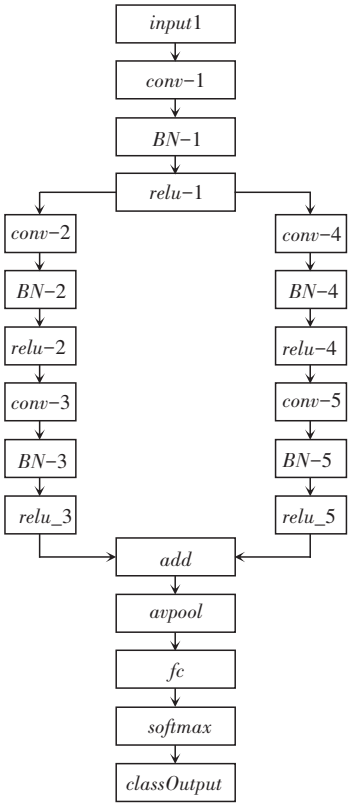

图 5 视线落点定位深度学习网络结构一

Fig. 5 The first type of network structure

表 2 网络结构中变量的释义

Tab. 2 Interpretation of variables in network structure

变量	含义
Input	输入的图像
conv − *i*(*i* = 1,2,3,4,5)	第 *i* 层卷积和多项式乘法
BN − *i*(*i* = 1,2,3,4,5)	第 *i* 层标准化
relu − *i*(*i* = 1,2,3,4,5)	第 *i* 层 ReLU 函数
add	所有特征值都纳入考量
avpool	平均池化
fc	全连接层
softmax	Softmax 函数
classOutput	类输出

(2)输入处理后的图像。由于第一种网络结构的训练效果并不理想,其中面部特征提取也不好,而且出现了重大偏差。尽管第一种设计的中间步骤有互相独立的 2 个分支分别进行了 2 轮卷积和多项式乘法、标准化等处理,但导致最终视线落点定位结果未臻至理想的原因可能是因为在初始阶段输入的图像过于庞大,在未能精准分辨面部位置情况下便把第一轮卷积和多项式乘法等处理的结果作为初始元送入后续处理。基于此,本文做出些许调整,在原本的网络结构不变的情况下将原本的 input1 换成 input2,input2 是处理后仅有人眼睛的图像,由此得到的处理后的测试结果如图 7 所示。

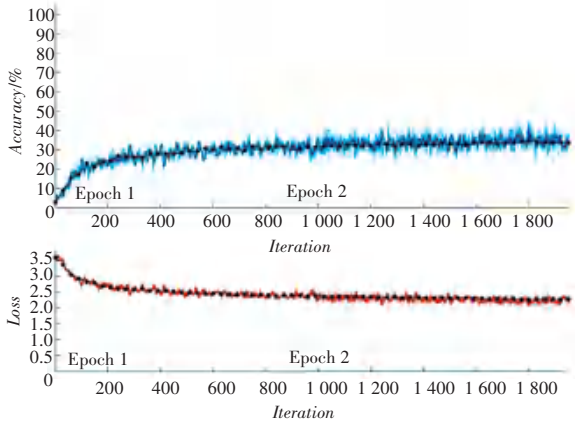

图 6 输入初始图像的训练结果

Fig. 6 Training results of inputting initial image

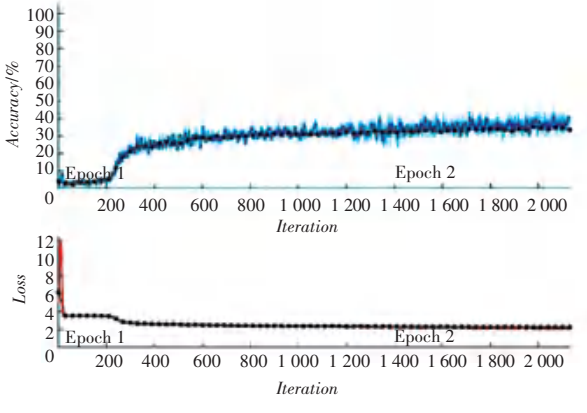

图 7 输入处理后图像的训练结果

Fig. 7 Training result of inputting processed image

蓝色线描述的采用处理后的图像的准确率最后只稳定在 33.44%,红色线训练集在模型中的预测结果与真实结果的误差也佐证这个网络设计的测试结果并不理想。通过计算所有 49 115 个预测点和原点的距离差及其平均值,即 2.045 个单位。鉴于本文所采用的屏幕仅有 5×7 个单位,相差 2.045 个单位的测试结果也较不理想。

2.3 视线落点定位深度学习网络结构二

由于前两次网络结构的训练效果并不理想。究其原因可知,第一次输入的是原图,背景中可能产生很多影响因素,导致面部识别产生偏差,进而使得视线落点定位出现重大偏差;第二次输入仅有眼睛的

图像，降低了面部识别误差的同时，却损失了人眼相对于面部的位置信息和面部相对于环境的位置信息。所以视线落点定位效果依然不够理想。综合考虑后将前文论述网络结构做出些许调整，在原本的仅有一个输入的情况下增加一个新的图像输入 input2，input1、input2 分别是原图和裁剪后仅有眼睛的图像。此外，为确保 input2 的特征提取不受 input1 的干扰，这 2 个图像分别各自进行了卷积和多项式乘法、标准化等处理，待特征值处理后再进行全连通深度学习。综合前述分析后可知，本文研究使用的第二种深度学习的网络结构如图 8 所示。由此得到的第二种网络结构的训练结果如图 9 所示。

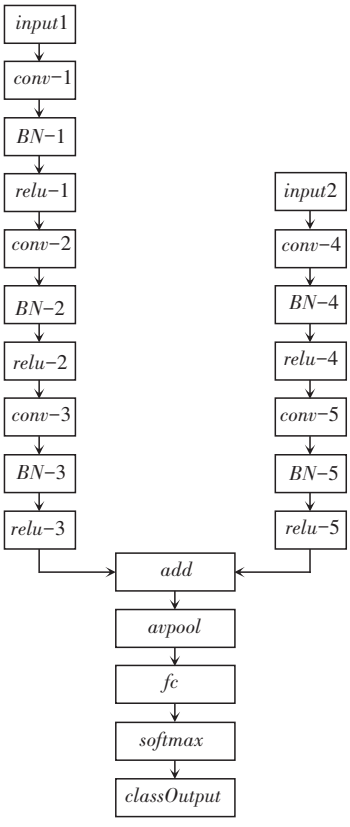

图 8　第二种网络结构图

Fig. 8　The second type of network structure

由图 9 可以清楚看到测试集的准确率能达到 49.60%，这个结果比前述可供对比的网络结构的准确率分别高 15.95%和 16.16%。而且图 9 的测试结果仅能得知预测的视线落定是否精准定位在测试区域，但无法得到通过深度学习预测的视线落点距离测试区域有多远。为此研究查看了训练后的预测数据，并计算了所有 49 115 个预测点和原点的距离差及其平均值，即 1.602 个单位。鉴于本文所采用的屏幕有 5×7 个单位，在未采用高精度仪器追踪视线的情况下，相差 1.602 个单位的测试结果较为理想。

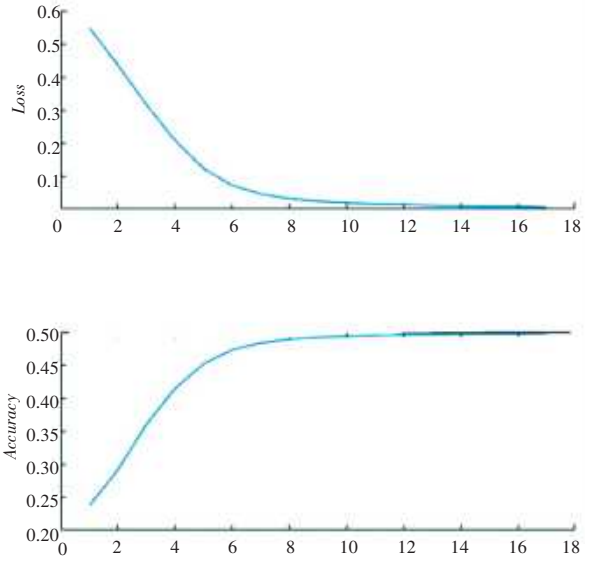

图 9 第二种网络结构的训练结果

Fig. 9　Training result of the second type of network structure

3　用户反馈流程

首先在某一分区投放一个产品如广告等，不妨假设在第 n 分区（具体位置见图 10）。用户反馈收集流程如图 11 所示。启动笔记本电脑的前置镜头拍摄画面，用迭代检测面部是否在画面中。若在这个画面中没有面部则返回上一步，即用前置镜头继续拍摄画面；若有面部存在，则用深度学习检测画面中人的视角落点区域。若该人的视角落点并未落在第 n 分区则返回上一步，即用深度学习检测画面中人的视角落点区域；若该人的视角落点位于第 n 分区，则识别该人的情感判断其人此时的情感，并计算其凝视第 n 分区的时长。把该人表现出的情感和凝视第 n 分区的时长作为用户对产品的反馈信息输出。

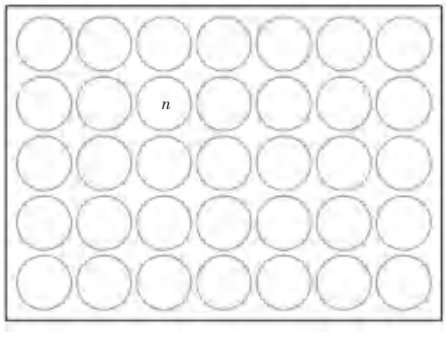

图 10　显示屏的 35 分区示意图

Fig. 10　35 partition schematic of the display screen

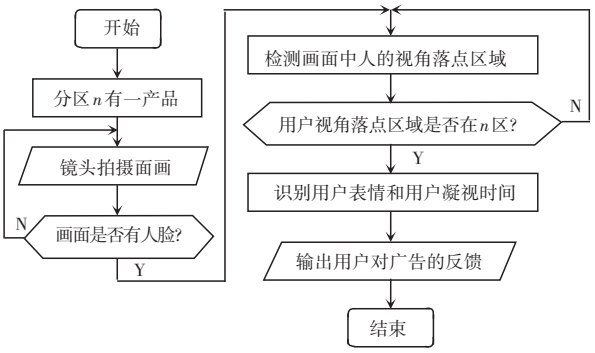

图 11 用户反馈收集流程图

Fig. 11 User feedback collection process

4 结束语

基于情感识别和视线追踪的用户反馈采集是一个极具发展潜力和商业价值的新兴课题。本文设计的研究创新主要可表述如下。

(1)通过对国内外相关文献资料的广泛调研及查阅,本文建立了一个可以实现面部情感识别的网站。

(2)在没有高精度仪器采集面部图像、且没有光学设备获得较为准确的眼动数据的情况下,通过深度学习实现准确率达 49.60%的视线追踪。

此外,面部情感识别和视线追踪技术均是多学科交叉的学界热点研究内容。其中,情感识别目前虽然已经陆续推出了很多不同的算法模型,取得了不错的识别效果,但却仍未能完全达到在实际环境中完美应用的要求。迄今为止,这也还是一个颇具挑战性的课题;而基于深度学习的视线追踪技术的视线落点定位准确率仍然偏低,故而亟需通过改善网络结构等方法提高视线落点定位准确率。期待本文工作能够为今后的深入探讨研究提供有益借鉴。

参考文献

[1] 高峰. 基于二维 Gabor 变换与支持向量机的人脸表情识别研究[D]. 天津:天津大学,2008.

[2] 施徐敢. 基于深度学习的人脸表情识别[D]. 杭州:浙江理工大学,2015.

[3] 邱玉. 基于动态表情识别的情感计算技术[D]. 宁波:宁波大学,2015.

[4] 程曦. 基于深度学习的情感识别方法研究[D]. 长春:长春工业大学,2017.

[5] 金辉,高文. 人脸面部混合表情识别系统[J]. 计算机学报,2000,23(6):602-608.

[6] 冯成志,沈模卫. 视线跟踪技术及其在人机交互中的应用[J]. 浙江大学学报(理学版), 2002, 29(2):225-232.

[7] KOTSIA I, ZAFEIRIOU S, PITAS L. Texture and shape information fusion for facial expression and facial action unit recognition [J]. Pattern Recognition, 2008,41(3):833-851.

[8] 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]//Proceedings of the 3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), CVPR 2010. San Francisco, CA, USA: IEEE, 2010: 94-101.

[9] HUANG Qiong, VEERARAGHAVAN A, SABHARWAL A. TabletGaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets[J]. Machine Vision and Applications, 2017, 28(5-6):1-17.

(上接第 62 页)

[2] HOLLIDAY A J, KAY J A. The use of infrared viewing systems in electrical control equipment [C]//Conference Record of 2005 Annual Pulp and Paper Industry Technical Conference. Jacksonville, FL, USA:IEEE,2005: 291-295.

[3] BAEK J, KIM J, YOON C, et al. Part-based hand detection using HOG[J]. Journal of the Korean Institute of Intelligent Systems, 2013,23(6): 551-557.

[4] LEE J P, LEE J Y,HYUN C H. Coin recognition and classification using digital image processing[J]. Journal of the Korean Institute of Intelligent Systems, 2012, 22(1): 7-11.

[5] LIN Qing, HAN Y J, HAHN H S. Lane detection in complex Environment using grid-based morphology and directional edge-link pairs [J]. Journal of the Korean Institute of Intelligent Systems, 2010, 20(6): 786-792.

[6] 彭智浩,杨风暴,王志社,等. 基于数学形态学和自动区域生长的红外目标提取[J]. 红外技术,2014,36(1):47-52.

[7] 徐青,范九伦. 新的基于分解直方图的三维 Otsu 分割算法[J]. 传感器与微系统,2017,36(1): 119-122,126.

[8] JAFFERY Z A, IRSHAD. Performance comparison of image segmentation techniques for infrared images[C]//12th IEEE India Conference (INDICON)-2015. Delhi:IEEE India Council, 2015:1-5.

[9] JAFFERY Z A, DUBEY A K. Design of early fault detection technique for electrical assets using infrared thermograms [J]. International Journal of Electrical Power& Energy Systems, 2014, 63: 753-759.

[10] FAN Songhai, YANG Shuhong, HE Pu, et al. Infrared electric image thresholding using two-dimensional fuzzy Renyi entropy [J]. Energy Procedia, 2011, 12:411-419.

[11] LI Ying, MAO Xingjin. An efficient method for target extraction of infrared images[C]//WANG F L, DENG H, GAO Y, et al. Artificial Intelligence and Computational Intelligence. AICI 2010. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 2010, 6319: 185-192.

[12] 吴龙国. 基于高光谱成像技术的土壤水盐及番茄植株水分诊断机理与模型研究[D]. 银川:宁夏大学,2017.

[13] 刘榴. 激光加工中视觉定位系统的研究[D]. 武汉:华中科技大学,2012.

[14] 周文欢,郑力新. 提取连通分量算法在棒材自动计数中的应用[J]. 微型机与应用,2011, 30(18):38-41.

[15] 刘军,李子毅. 一种复杂背景环境下的改进型 PCNN 图像分割算法[J]. 计算机与数字工程,2018,46(2):375-381,406.