Non-contact Monitoring of Body Parameters during Human-machine Interactions for the Assessment of Affective Health

Thesis submitted to
Indian Institute of Technology Kharagpur
for the award of the degree

of

Doctor of Philosophy

by

Satyajit Nayak

under the supervision of

Monalisa Sarma
Assistant Professor
(Subir Chowdhury School of Quality and Reliability)
Aurobinda Routray
Professor
(Department of Electrical Engineering)

SUBIR CHOWDHURY SCHOOL OF QUALITY AND RELIABILITY
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

August 2022

Abstract

Affect is a subconscious experience of the intensity of emotion. The affective states play a central role in everyday decisions and well-being. Developing systems for recognizing and managing an individual's affect become an integral domain in Human-machine Interaction (HMI) research. There is an increasing focus on developing emotional artificial intelligence in HMIs. The human affect assessment can be identified using verbal (explicit) or non-verbal (implicit) cues. However, affect recognition during HMI addresses several challenges such as: natural setup, a comprehensive collection of potential humans' interactions, mapping from genuine emotion to affective states, the obtrusiveness of the sensing equipment affects the user response to emotion triggering, the lack of observability in real-life settings, and inaccurate ground-truth affect labels. HMI researchers employ specific technological interventions to assess human affective health using different modalities during machine interactions. One of the prime concerns is the lack of standard non-contact databases for developing and evaluating affective health. Due to certain limitations of the state-of-the-art databases motivated me to build and validate with proper ground truth of our in-house created database named CARD: Contactless Affect Recognition Database adopting Synchronize Multi-modality. It consists of visual images, shorter visual and thermal clips, infrared or thermal facial blood flow rate, breath waveform, and heart waveform of a person while interacting with a made to design Health Kiosk. Auto identification of affective health such as stress, anxiety, and depression levels in human beings while interacting with computers or machines has become essential for preventing mental health issues. Visual facial expressions form an essential cue that satisfies these criteria. An unobtrusive solution for estimating stress level through a metric based on the fraction of the time an individual exhibits stressed facial expressions. We developed a real-time portable embedded device named StrexNet, consisting of the required hardware and a customized App module to recommend the steps for

perceived stress control at the user end. The classification of facial expressions has been carried out using a light Convolutional Neural Network (CNN) architecture. The network is short and well suited to work with embedded processors. We have validated the metric using t-test and two-way ANOVA with p-values of 0.047 and 0.036, respectively, based on Depression Anxiety Stress Scale (DASS-21) self-evaluation scores on stress. An accuracy of 94.58% is obtained while comparative analysis with state-of-art literature is performed.

Thermal or infrared facial imaging is also a non-contact mechanism for assessing the affective health of an individual during HMI. Real-time detection of the face and tracking the Regions of Interest (ROI) in the thermal video during HMI is challenging due to head motion artifacts. We proposed a three-stage HMI framework for computing the multivariate time-series thermal video sequences to recognize human emotion and provide distraction suggestions in real-time. The first stage comprises a face, eye, and nose detection using a Faster R-CNN (region-based convolutional neural network) architecture and used Multiple Instance Learning (MIL) algorithm to track the face ROIs across the thermal video. The mean intensity of ROIs is calculated, forming a multivariate time series (MTS) data. In the second stage, the smoothed MTS data are passed to the Dynamic Time Warping (DTW) algorithm to classify emotional states elicited by video stimulus. Our proposed framework provides relevant suggestions from a psychological and physical distraction perspective during HCI in the third stage. Our proposed approach signifies better accuracy than other classification methods and thermal data sets. For better precision in affect recognition, twin channels of thermal and visual facial image sequences to register an individual's affective states are up-and-coming compared to different cues such as visual and thermal facial images. A two-stage smart engine is proposed for classifying the dominant-negative affective state by clustering sequences of emotional states. The first stage obtains the dominant emotional state by an ensemble of cues from visual and thermal face images using a newly proposed cascaded Convolutional Neural Network (CCNN) model as it classifies the dominant emotional state of an individual. The second stage clusters a sequence of the obtained emotional states using a trained Hidden Markov model (HMM) as one of the four dominant affect - stress, depression, anxiety, or healthy. The performance second stage has been compared with a standard DASS score.

We also developed a non-contact framework for estimating shortness of breath patterns due to the negative psychological states of an individual while interacting with the computer. And recommend the relevant suggestions to overcome the shortness of breathing through a customized android application with the help of a registered healthcare expert. The framework incorporates an FMCW Doppler radar sensor that operates at 40 GHz can supply HR, BR, heart, and breath waveform with a sampling rate of 20 frames per second, along with a k-nearest neighbor breath classification algorithm. By calculating the threshold inhalation Tidal Volume (TV), exhalation TV, BR, and HR of 60 seconds, the proposed breath classification algorithm categorizes the breath waveform into shortness, normal or conscious, and deep breathing. Shortness of breath classification validated with the University of California, San Diego (UCSD) subjective evaluation score. Doppler radar breath signal is calibrated with 32-bit polysomnography (PSG) channel device on contact mode using the Pearson correlation coefficient between two signals. The correlation coefficient of signals for shortness of breathing is a maximum of 0.97, and the average is 0.75, respectively. We studied the physiological health signatures using 32-bit(PSG) across the 45 healthy participants while watching the negative psychological stimulus.

Keywords: Affective Health, Facial Expression, Facial Blood-flow Rate, Shortness of Breath, Human-machine Interaction, Portable Embedded Systems.