Unobtrusive Academic Emotion Recognition Based on Facial Expression Using RGB-D Camera Using Adaptive-Network-Based Fuzzy Inference System (ANFIS)

James Purnama, Swiss German University, Indonesia

Riri Fitri Sari, University of Indonesia, Depok, Indonesia

ABSTRACT

Quality of learning in the classroom is influenced by many factors. One of them is the academic emotions of the students. The emotion detection in the classroom cannot be done by using sensors attached to the body of the students, because it would disturb the concentration of the students. The proposed solution is by using unobtrusive emotion detection, e.g. by placing video capture equipment, which is not visible at the front of the student's desk. In this study, an RGB - Depth Microsoft Kinect camera is used to record facial expressions by considering the convenience factor of the students, speed of response time, and cost efficiency. A combination of Cohn-Kanade dataset and EURECOM dataset is used as the training set in machine learning with Adaptive-Network-Based Fuzzy Inference System (ANFIS) algorithm, with 8 sample of Asian race students (4 male and 4 female students).

KEYWORDS

Academic Emotion Detection, ANFIS, Facial Expression, RGB-D Camera

INTRODUCTION

In the era of Information Technology, the teaching and learning process in the classroom is still the main method of learning. In addition, there are e-learning methods and Intelligent Tutoring System, which are supplements to the classroom learning method. Quality of learning in the classroom is influenced by many aspects, one of which is the Academic Emotions.

Pekrun et al. (2002) describes that the Academic Emotions can affect learning qualities which confidence, excitement, frustration, interest, flow/engagement, boredom, confusion, and anxiety (Arroyo et al., 2009; Azcarraga et al., 2011; Burleson, 2006; Azcarraga et al., 2010; Kapoor et al., 2007; Zeidner, 2007).

Previously, emotion detection used sensor devices that are plugged / affixed to the body of the research subject. The sensor devices are physiological signals or biosignal sensors to measure body signals, such as Electromyography (EMG), Electro Dermal Activity (EDA) or Galvanic Skin Response (GSR). This method is commonly referred to as the Hidden Ways (covertly). This method was perceived only suitable for testing in the laboratory, and are not effective when implemented in

real conditions, since mounting the sensor devices obviously disturb the concentration of the research subjects, and are likely to affect the feelings / emotions (Guo et al., 2013).

The concept of unobtrusive emotion detection is to ensure that the detection process is not flashy or attracts undue attention from the research subjects. The main objective of our work is to create the unobtrusive environment which give comfort to the user, by working as much as possible so that the user does not aware of the existence of the sensor device used (Broek et al., 2009).

Haq and Jackson (2011) concluded that unobtrusive emotion detection analysis is a combination of a multimodal channel, such as audio (speech expressions), and video (facial expression and body gesture). Reisenzein et al. (2013) proved that the relation between facial expression and emotion is coherence. Based on their experiments, we know that emotion such as "amusement" and "smiling" have high coherence, while other positive emotions and smiling have low to coherence relation. Jack et al. (2012) published revolutionary findings, that facial expressions of emotion are not culturally universal. It means the Western Caucasian people have different facial expressions compared to Eastern Asian to express their emotion. These findings give an impact and is challenging to the researchers of emotion detection.

This research focuses on facial expressions-based emotion detection which is mainly in posed facial expressions recognition. Since the last few years, the research direction has changed to the spontaneous expression (Bettadapura, 2012). This new direction is due to the rise of low-cost and powerful RGB-D camera sensor, such as Microsoft Kinect (Zhang, 2012; Andersen et al., 2012) and other sensors such as the PlayStation Eye (Sonny, 2014).

RELATED WORKS

Research in emotion detection has had a lot of interest since decades ago, marked by a lot of publications in this area. With the birth of technology innovation of RGB-D sensors (depth camera) which are cheap and affordable during the last few years ago, research in the field of emotion detection start a new chapter, namely unobtrusive emotion detection.

Mahmoud et al. (2011) has described the collection and annotation of a 3D multi-modal corpus of naturalistic complex mental states. The corpus consists of 108 videos and 12 mental states collected from a Microsoft Windows Kinect depth camera. The main contributions of their research are that they provide multimodal 3D corpus of the combination of facial expressions, as well as hand and body gesture.

Lucey et al. (2010) have added a few features on the preexisting Cohn-Kanade dataset (Kanade et al., 2000). They extended the famous Cohn-Kanade facial expressions dataset with emotion detection features using Active Appearance Models (AAMs) and a linear Support Vector Machine (SVM) classifier. This famous dataset is called the Extended Cohn-Kanade (CK+). The accuracy of CK+ was compared with the Karolinska Directed Emotional Faces (KDEF) dataset, and the result was described in Table 1 (Lucey et al., 2010). Their experiments showed a good result in predicting emotion with the average accuracy for 7 emotions predictions of 74.57%.

Huynh et al. (2013) used a Microsoft Kinect depth camera sensor to detect facial expressions. The main objective of their research is to recognize gender by introducing a novel efficient Local Binary Patterns (LBP) algorithm based on descriptor, namely Gradient-LBP. The set of depth image data captured from Microsoft Kinect depth camera were formulated into a dataset named EURECOM, and processed along with Texas 3D Face Recognition Database (3DFRD) (Gupta et al., 2010).

Xia et al. (2011) optimized the usage of Microsoft Kinect for human detection. They proposed a novel approach by using 2-D head contour and model 3-D head surface model. These approaches were something new compared to the existing method that used features based on gradients, such as histograms of oriented gradients (HOG), or extract interest points in the image, such as scale-invariant feature transform (SIFT), etc. The accuracy rate of their method was 98.4%.

Related Content

Learner-Centered Pedagogies: A Critical Review of the Use and Implications for Learner-Centered Pedagogies
Anne W. Kanga (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications (pp. 1559-1576).*
www.irma-international.org/chapter/learner-centered-pedagogies/237951

A Generic Framework for Feature Representations in Image Categorization Tasks
Adam Csapo, Barna Resko, Morten Lind, Peter Baranyiand Domonkos Tikk (2012). *Software and Intelligent Sciences: New Transdisciplinary Findings (pp. 491-512).*
www.irma-international.org/chapter/generic-framework-feature-representations-image/65147

A Robust Intelligent System for Detecting Tomato Crop Diseases Using Deep Learning
Mohamed Afify, Mohamed Loeyand Ahmed Elsawy (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-21).*
www.irma-international.org/article/a-robust-intelligent-system-for-detecting-tomato-crop-diseases-using-deep-learning/304439

Formal Methods for Specifying and Analyzing Complex Software Systems
Xudong He, Huiqun Yuand Yi Deng (2007). *Advances in Machine Learning Applications in Software Engineering (pp. 319-345).*
www.irma-international.org/chapter/formal-methods-specifying-analyzing-complex/4866

Person Identification System in a Platform for Enabling Interaction With
Individuals Affected by Profound and Multiple Learning Disabilities