3D Face Tracking in Emotion Recognition

A specific challenge that has been particularly highlighted in the 21st century is the need to improve human-computer interfaces because of the increased integration of technology in almost all activities. Recently, artificial intelligence has been highly regarded as a potential way of making computers intelligent and improving the human-computer interfaces. As a result, there is remarkable achievements in integration of computer visual technology in various fields including biometrics, marketing, surveillance, image search, teleconferencing, and indexing among other fields. The successful integration of computer visuals in the aforementioned fields has led to massive improvements in security, fraud avoidance, and reduction of various technology-related crimes. Despite current use of 3D face tracking in various fields, there is still a knowledge gap in using facial recognition technology to perform emotional analysis using expression recognition, mental state estimation or intention prediction.

Emotion recognition and intention analysis are natural capabilities found in a limited group of living things. However, recent technological developments have been targeted towards developing systems that are able to replicate the entirety of human’s existence right from cognitive ability to understanding emotions, feelings, and intentions (Smith 54). Despite existing humanoids and robots doing almost everything that a normal human being can do, one significant challenge has been developing emotion recognition systems that are able to construe accurate information by performing emotional analysis and understanding a person’s state of mind (Dong 1157). Any computer that can possibly recognize human emotion and intentions could substantially solve all current human-computer interaction (HCI) problems. It could also significantly change the way in which different human activities are conducted and perceived. Therefore, there is ongoing research aimed at creating a paradigm shift from human-computer interaction to human-human interaction (Davis 3). As a result, different fields such as healthcare will have the opportunity to create fulfilling, natural, and productive relationships between people and technology.

It is generally accepted that the human face can display emotions and can be used to identify and discriminate people based on their facial movements, lip movements, and a myriad of other facial identifiers. After understanding this aspect of the human face, it is no longer as surprising that most emotion recognition systems are targeting at replicating and modelling this feature. According to a study by Ekman and Friesen, facial expressions in human beings communicate more than half of the intended message, facial intonations communicate about 40%, and less than 10% is through mere words (Ekman & Friesen 154). From that study, it is clear than facial expressions are critical aspects of human communication. Moreover, it is hypothesized that that all six primary human emotions of sadness, happiness, anger, surprise, disgust, and fear exhibit unique facial expressions that are universal in all human beings. Understanding that aspect of the human face provides an objective basis of developing emotion recognition frameworks.

Towards documenting human emotion, Ekman and Friesen developed a facial action coding system (FACS) that was aimed at measuring facial activity by providing a linguistic description of detectable facial movements (Ekman & Friesen 163). The target of this coding system was to develop a standard way of documenting human expression by eliminating biases introduced by errors of human coding. Since then, there have been developments of other emotional facial coding systems that are focused on using standard actionable units to represent human emotion. However, in the year 1994 there were criticism of universal coding of human emotions citing specific factors such cultural differences, skin color, head movements and aspects of muscle movement as significant inhibitors. Later in the same year, Ekman and Friesen replied to the criticism by providing standard algorithms and systematic point-by-point arguments aimed at providing evidence that human facial expressions are universal and consistent across people of all cultures. Despite substantial developments in emotion recognition to date, it is still widely debated how different facial kinetics at different situations accurately represent human intentions and emotions.

Other frameworks were developed over time to detect the human face and develop landmarks that could facilitate capturing, analysis, and interpretation of emotion. A research analysis proposed a unified tree-structured model that could be used to model poses, facial detection, and estimation of facial landmarks (Dong et al., 1157). This approach was encoded-based and was highly dependent on changes of viewpoint and each possible facial landmark was represented as a dynamic part model. In that regard, a dynamic coding algorithm was constantly used to develop realistic solutions. One of the biggest challenges in this conceptual framework was the lack of an effective way of eliminating facial landmarks. To further improve this approach, it was suggested that the use of specific part mixtures that eliminate intermediate landmarks by introducing part mixtures by highlighting and marking the most important facial anchor points. Despite the efforts put in developing this model, there was a huge challenge in developing an algorithm that would accurately represent human emotion.

Recently, Discriminative shape regression (DSR), system has been used to eliminate misaligned facial landmarks by accurately and intermediately locating facial landmarks across different people’s faces under different illuminations and viewpoints (Dong et al. 1157) This face tracking technique relies on extraction of uniformly distributed random points on the human face which are often generated in a cascaded 2D form. Before they are generated into 2D form, this system performs an initial guess of the human shape then progressively refines that guess using different feature spaces. It is important to indicate that these feature spaces are regressed and can be affected by differences in pixels or any other binary features developed from the face with respect to landmarks. This regression model proves to be better than previous coding systems because it is real-time and relatively representative. Eventually these cascaded 2D forms are used to extract and develop facial features that systematically identify unique attributes in different faces.

After the facial features are systematically developed, the next challenge is developing mathematical models aimed at interpreting facial data in the attempt to deduce emotion, intention, or any other content. Principal component analysis (PCA) is one technique that is applied in modelling variance of features in the human face with the aim of developing a component of universality (Smolyanskiy 865). However, this stage is prone to many errors because most of the computational information required is greatly affected by cluttered backgrounds, viewpoints, and illuminations that hinder the process of developing accurate models. Real-time 3D tracking has also been focused on finding automatic initialization and reacquisition of images with the aim of achieving significant equations that are not highly affected by factors such as occlusion and instability. One system that has been used recently do so effectively is the use of a pose correction mechanism that dynamically records randomly distributed facial points with the aim of generating a consistent range. By doing so, an accumulated error is measured which allows mathematical models to effectively mitigate variability and other limiting effects. However, this technique is not successful in providing human emotion because of biased tracking points and image reacquisition challenges.

Also, a recent study suggests the use of an active appearance model (AAM) that tries to resolve impending challenges in visual technology by constricting the dimensions of the human face in a specific linear 3D morphable model (Smolyanskiy 867). Consequentially, AAM algorithms are developed using developed landmarks and regressive face models are subsequently tracked using a 3D morphable mechanism. Because of lack of accuracy and robustness in this model, a RGBD camera is introduced to reduce tracking errors by using visual-based tracers to capture and interpret depth data. Despite this approach deducing a method of reducing facial occlusions, it fails when more than a quarter of the face is occluded and specific facial features such as thick glasses and facial hair can significantly affect accuracy. Again, this approach was vastly unsuccessful in performing emotion analysis because of the lack of a standard way of representation.

Future work in using 3D face tracking to capture and interpret human emotion is on developing specific algorithms aimed correlating the specific facial expressions with the pixels that display the highest model performance. To do so effectively, the number of facial muscles engaged in every facial movement will need to be identified and differentiated for each facial expression (Fernandez-Dols 33). For instance, it can be inferred that emotions that require minimal activity of face muscles will have smaller pixel dimensions and emotions which require more facial muscle activity will be have larger pixel dimensions. As a result, further knowledge on range of human expressions will be easily captured, interpreted, and communicated. However, there is a huge challenge in understanding the intensity of human expressions and more research will need to be done because emotions such as anger differ from one person.

On that basis, it is evident that 3D facial recognition remains one of the most complex artificial intelligence developments in the 21st century despite a lot of research being directed towards improving it. After analysis of progression of facial recognition technology through the years, it is highly plausible that future 3D facial tracking systems will accurately perform emotional analysis to give a full range and intensity of human emotions, intentions, and states of mind.

Works Cited

Davis, Ernest. "Algorithms and everyday life." *Artificial Intelligence*, vol. 239, 2016, pp. 1-6.

Dong, Yanchao, et al. "Real Time 3D Facial Movement Tracking Using a Monocular Camera." *Sensors*, vol. 16, no. 8, 2016, p. 1157.

Ekman, P., and W. V. Friesen. "What Are the Similarities and Differences in Facial Behavior across Cultures?" *Emotion in the Human Face*, 1972, pp. 153-167, doi:10.1016/b978-0-08-016643-8.50030-6.

Fernández-Dols, José-Miguel, and Carlos Crivelli. "Recognition of Facial Expressions: Past, Present, and Future Challenges." *Understanding Facial Expressions in Communication*, 2014, pp. 19-40.

Smith, Marcus, et al. "Facial recognition." *Biometrics, Crime and Security*, 2018, pp. 54-70.

Smolyanskiy, Nikolai, et al. "Real-time 3D face tracking based on active appearance model constrained by depth data." *Image and Vision Computing*, vol. 32, no. 11, 2014, pp. 860-869.