EasyChair Preprint

№ 6907

Survey on Upper Body Gesture with Facial Emotions Recognition

Ravi Shankar Sharma and Pratima Gautam

October 20, 2021

Survey on upper Body Gesture with Facial Emotions Recognition

Ravi Shankar Sharma, Dr. Pratima Gautam

Rabindranath Tagore University, Bhopal,

mca.ravi2011@gmail.com

Abstract:

Real Time emotion recognition and Body Gesture has become a trending research topic. While works based on facialexpressions with upper body Gesture recognizing a less explored topic. We present a new comprehensivesurvey hoping to enhance research in the field. We first introduce upper bodyActivity as a component of what is commonly knownas "Human body language". We then Compare a completeframework for automatic emotional body gesture recognition. We also compare person detection andbodypose estimation methods both in 2D(RGB) and 3D. We then comment the recent literature related to representation learning and emotionrecognition from images of emotionally expressive gestures. We also discuss multi-modal approaches that combine facewith body gestures for improved emotion recognition. We then Analysis Existing Algorithm for Human Activity and Emotion's and Compare Real Time Facial Emotion from Facial Expression Using Deep Learning with Genetic Algorithm.While pre-processing methodologies (e.g. human detection and pose estimation)are nowadays mature technologies fully developed for robust large scale analysis, we show that for emotion recognition the quantity oflabelled data is scarce.

Keywords:body language, upper body gesture, emotion recognition, body pose estimation, affective computing, Deep Learning, Genetic Algorithm.

1. INTRODUCTION

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do. Computer Vision, often abbreviated as CV, is defined as a field of study that seeks to develop techniques to help computers "see" and understand the content of digital images such as photographs and videos.

Automatic facial expression recognition system has many applications including, but not limited to, human behaviour understanding, detection of mental disorders, and synthetic human expressions. Two popular methods utilized mostly in the literature for the automatic FER systems are based on geometry and appearance. Even though there is lots of research using static images, the research is still going on for the development of new methods which would be quite easy in computation and would have less memory usage as compared to previous methods.

2. Literature Review

Juan Wu etc. [1] in his Research paper "Research on Computer Vision-Based Object Detection and Classification" describe that they are explicitly explore current advances in the field of object detecting and categorizing based on computer vision, and a comparison of these methods is given.For Object Detection we use Color Based Approaches and work with Color Index We divide shaped-based approaches into two categories: boundary-based approachesand region-based approachesexistingobject detection methods associated with Color or shape.

Xin jia [2] in his Research paper "Image Recognition Method Based on Deep Learning" Describe that its discuss Deep Learning Algorithm basically divided into four parts -Convolutional NeuralNetworks, Restricted BoltzmannMachines, Autoencoderand Sparse Coding.For theapplications in the computer vision domain, the papermainly reports the advancements of CNN based schemes,as it is the most extensively utilized and most suitable for
images. Most notably, some recent articles have reportedinspiring advances showing that some CNN-basedalgorithms have already exceeded the accuracy of humanraters.Despite the promising results reported so far, there

is significant room for further advances.

Federica Scarpina et al. [2] proposed the result might agree with the hypothesis about affected individuals' difficulties in being attentive to negative facial emotions, and specifically in the case of fearful expression. This study might encourage future research in which emotional processing will be investigated through subjective judgments and implicit/objective measurements.

ElżbietaKukla et al. [17] proposed investigations will concentrate on separate recognition of the two parts of faces that are the most important in emotion identification, i.e. regions of eyes, eyebrows and forehead (upper parts) and regions of mouth, nose and chin (lower parts). Particular networks could be trained and used to recognize separately particular action units from FACS coding system. Final recognition results are this case could be a fusion of the results obtained by individual classifiers.

Hsi-Chieh Lee et al. [18] proposed a system to automatically recognize the facial expressions. Our system extracted and described the features from the contour of eyebrows, eyes and mouth by a scalable rectangle. This is an improvement over the ways for using manual facial characteristic points and complicated face mask model. We defined less features to reduce the recognition time and obtain appropriate recognition accuracy.

DjamilaRomaissaBeddiar et al. [3] proposed to understand and interpret effectively human activities has become unavoidable in several applications of computer vision, HCI, robotics, security and home monitoring. This paper aims to give an overview of the recent works in this field of research. It proposes a classification according to several criteria. It initially discusses the different applications of HAR, and the major objectives intended by these systems. Then, it presents an analysis of the used approaches in the state of the art, as well as the means used in their validation.

Shugang Zhang et al. [15] proposedhuman activities into three levels including action primitives, actions/activities, and interactions. We have summarized the

classic and representative approaches to activity representation and classification, as well as some benchmark datasets in different levels. For representation approaches, we roughly sorted out the

research trajectory from global representations to local representations

and recent depth-based representations.

Daniel Canedo et al. [7] proposed researchers became aware of the potential in CNNs for solving Computer Vision problems, and more FER systems using CNNs emerged, correlating with overall better results. The only potential negative aspect to point out from the reviewed works is that none considered the environment context. Although most works are giving the right steps towards multimodal systems, the environment context seems to be ignored. For instance, if there is an image of a birthday party, the happy context has a huge weight in the mood of people participating in it, which can't be ignored even if a certain participant is not explicitly smiling. Nevertheless, FER systems are being stimulated by yearly challenges and by the overall interest in numerous fields, achieving better results year by year.

Yi Liu et al. [9] proposed on the state-of-the-art progress of 2-D human pose estimation based on deep learning. Classification, introduction, and performance comparison of the methods are involved. such as gradient decent and parameters mounting, need to be solved. Multi-branch CNN methods synthesize results of multiple CNN branches, at the cost of complex calculation and training. RNN methods applies the sequential processing used in Natural Language Processing to model the interrelationships between body joints, expands the receptive field with iteration of the recurrent network, and solves the occlusion problem in a certain degree at a cost of massive calculation. GAN methods capture the high order correlation of the data in the absence of the target class label information. Such methods have provided the state of- the-art best solution against the occlusion. Nonetheless, the drawback is still the complex calculation and the time cost.

Lisa Schrader, et al. [5] proposed machine learning methods from the field of human activity recognition (HAR) to detect human activities. These algorithmic methods need a large database with structured datasets that contain human activities. Compared to existing data recording procedures for creating HAR

datasets, we present a novel approach, since our target group comprises of elderly and diseased people, who do not possess the

same physical condition as young and healthy persons.

References

1. Juan Wu1, Bo Peng1, Zhenxiang Huang1, and Jietao Xie2 “Research on Computer Vision-Based Object Detection and Classification” IFIP International Federation for Information Processing 2013.

2. Xin Jia “Image Recognition Method Based on Deep Learning” 29th Chinese Control And Decision Conference (CCDC) 2017.

3.Lisa Schrader, et al., “Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People”, Published online Springer 24 February 2020.

4.Federica Scarpina, Giorgia Varallo, Gianluca Castelnuovo, Paolo Capodaglio, Enrico Molinari, Mauro Alessandro “Implicit facial emotion recognition of fear and anger in obesity”, Springer publication September 18, 2020

5.Yi Liu, Ying Xu, Shao-Bin Li," 2-D Human Pose Estimation from Images Based on Deep Learning : A Review", 2nd Advanced Information Management, Communicates, Electronic and Automation Control Conference, IEEE 2018

6.Daniel Canedo, and António J. R. Neves, "Facial Expression Recognition Using Computer Vision: A Systematic Review", www.mdpi.com/journal/applsci , Published: 2 November 2019

7.Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li, "A Review on Human Activity Recognition Using Vision-Based Method", Hindawi Journal of Healthcare Engineering Volume, Article ID 3090343, 31 pages https://doi.org/10.1155/2017/3090343, 2017.

8.Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou, Abdenour Hadid, "Vision-based human activity recognition: a survey", Published online Springer 15 August 2020

9.Hsi-Chieh Lee1, Chia-Ying Wu, and Tzu-Miao Lin, "Facial Expression Recognition Using Image Processing Techniques and Neural Networks", *Advances in Intelligent Systems & Applications*, SIST 21, pp. 259–267, Springer-Verlag Berlin Heidelberg 2013.

10. Elżbieta Kukla, "Facial Emotion Recognition Based on Cascade of Neural Networks", Springer International Publishing Switzerland 2015