Name: URVI PATEL UID:22BIT034
Subject: SMM Date: 15/2/2024

A Literature Review: Emotion Recognition in Computer Vision Surveillance

Papers:

1. ‘Affective’ computing and emotion recognition systems: The future of biometric surveillance? Joseph Bullington
2. Emotion Recognition in Context: Ronak Kosti, Jose M. Alvarez, Adria Recasens, Agata Lapedriza
3. A study on computer vision for facial emotion recognition: Zi-Yu Huang, Chia-Chin Chiang, Jian-Hao Chen, Yi-Chian Chen, Hsin-Lung Chung, Yu-Ping Cai & Hsiu-Chuan Hsu
4. A Novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly: lZixiang Fei, Erfu Yang, Leijian Yu, Xia Li, Huiyu Zhou, Wenju Zhou

Vocabulary:

ubiquitous computing, probe, gallery, affect recognition, surveillance, biometrics, privacy, groupthink, multimodal, kinesthetic, human-computer interaction, emotion recognition, context, EMOTIC, clustering, Fleiss’ Kappa, truncated low rank filtering, VAD, SoftMax, Fusion Network, Feature extraction module, Jaccard Index, Validation set, Test Set, Stochastic Gradient with Momentum, CNN, regression model, Euclidean Loss, Global Average Pooling layer, first fully connected layer, Rectifier Loss Units, Average Precision, FER, Ekman, ResNet, vanishing gradient, window based cross attention mechanism, multi-scale features, regularization, class activation mapping, attention maps, self-attention mechanism, normalization, DNN, SENet, visual transformer, correlation, feature channel, data augmentation, epoch, normalized confusion matrix, ensemble CNN.

Abstract:

Emotional AI can be integrated in surveillance- for use cases involving drowsiness detection, groupthink prevention and insider-threat detection. It explores how consented analysis of our facial expression is the future of society, it being implemented in China and by recruiting companies already. An EMOTIC database is designed and is used with CNN to train a model to predict emotions taking context as a factor too. The context is provided by annotating these images in 26 distinct categories and 3 continuous ones. The paper emphasizes that such a technology holds potential in fields of human-computer interactions like AI assistance or online education. A CNN performs FER using major information conveyed by facial regions. It makes use of a cross-database validations and talks about on importance of mouth, nose and large patches of distinctive facial areas for model to accurately learn and predict emotions. SVM and MobileNet are used to develop a convolutional network which helps detect cognitive impairments through emotions detected while watching a stimulating video.

Introduction

Emotion Recognition, a derivative of Computer Vision and subsequently one of Deep Learning, analyses multimodal data to detect underlying sentiments by neural network technology and image-feature extraction. Deep Learning is faster in training models on extensive data, and hence renders itself suitable for applications requiring quick processing like surveillance. Computer Vision, in itself a growing field, teaches machines to understand the presence of objects, people and other targets and derive information from them. When trained in context, neural networks can better catch the suitable underlying feeling, which may or may not be overtly expressed in the image. Using this mix of computer vision and emotion analysis in surveilling citizens is a technology currently used only in Shanghai. However, beyond the Orwellian aspect of monitoring citizens' pulse, persisting emotions are worthy indicators of cognitive impairments. Emotion recognition can be one of the technologies used to flag these conditions. This is much like Computer Vision being used for detecting malaria by distinguishing parasitic cells. Through this Literature Review, I have studied 4 papers which helped me understand the possible use cases of real-time emotion recognition. They explore the creation of datasets, keeping diversity and context in mind, while examining which neural network architecture is giving the best results. Lastly, I have studied how we can use emotion analysis in videos to predict the presence of cognitive impairment in older adults.

Summary:

The paper **"Affective Computing and Emotion Recognition Systems: The Future of Biometric Surveillance?"** by Joseph Bullington, September 2005, discusses the implications of ubiquitous computing and the extensive use of surveillance cameras in England, Scotland, and EU countries. The author dismisses the debate on the effectiveness of surveillance cameras in reducing crime rates versus collecting citizen data, emphasizing the inefficiency of human surveillance and the potential of biometric technologies, particularly facial recognition, to address this challenge. The paper introduces the concept of affective computing and emotion recognition in the context of surveillance, highlighting the potential of real-time multi-modal input from crowded urban environments to accurately infer individuals' motives based on their underlying emotions. The author also addresses the privacy concerns and ethical implications associated with the use of affect recognition in surveillance, presenting various scenarios where emotional recognition could be applied, such as in transportation settings and group decision support systems. The paper concludes by calling for discussions on the implications of emotion AI surveillance in the security community before the widespread implementation of such technology. It also discusses the potential applications of affect recognition in detecting cognitive impairment in elderly individuals and the ethical considerations associated with its implementation.

The study **"Emotion Recognition in Context"** by Ronak Kosti, Jose M. Alvarez, Adria Recasens, and Agata Lapedriza, emphasizes the importance of considering context in classifying human expressions. The main finding of the study is that integrating body and image information is crucial for accurate emotion recognition. The authors developed the EMOTIC database, which includes images of people displaying various emotions along with the situation they are in. They used a combined loss function (Lcomb) with different Convolutional Neural Network (CNN) architectures and found that the highest Average Precision (AP) scores are consistently obtained when both body and image information are used as inputs, except for the "esteem" category. This suggests that, for discrete category recognition, considering both information sources provide the best results. The study also observed that the results using only the image are generally worse compared to using the body or both, implying that while the image contributes to emotion recognition, it alone doesn't offer sufficient information for accurate recognition. Therefore, the study's main finding is that integrating body and image information is crucial for accurate emotion recognition.

The study **"A study on computer vision for facial emotion recognition"** by Zi-Yu Huang, Chia-Chin Chiang, Jian-Hao Chen, Yi-Chian Chen, Hsin-Lung Chung, Yu-Ping Cai & Hsiu-Chuan Hsu, published in Scientific Reports, 2023, explores the use of Convolutional Neural Networks (CNN) for Facial Emotion Recognition (FER) and emphasizes the importance of considering data diversity and transfer learning in developing and deploying FER algorithms.

The study uses feature maps to identify important facial features for emotion classification, with a focus on the mouth, nose, and large patches of distinctive facial areas. The accuracy of emotion prediction is influenced by the number of images in the dataset, with higher accuracy for emotions like happiness and lower accuracy for emotions like disgust due to the availability of images. Transfer learning from the AffectNet dataset to the RAF-DB dataset improves the classification accuracy, but both models struggle with classifying neutrality. The SE-ResNet architecture is used for training and testing, and cross-database validation is performed to evaluate the model's performance. The study concludes that models trained on AffectNet show lower performance on other datasets, but perform better on RAF-DB, which is more constrained in terms of ethnicity and image range.

The paper titled **"Deep Convolution Network Based Emotion Analysis for Automatic Detection of Mild Cognitive Impairment in the Elderly"** introduces a novel approach to identify mild cognitive impairment (MCI) in the elderly population. Authored by Zixiang Fei, Erfu Yang, Leijian Yu, Xia Li, Huiyu Zhou, and Wenju Zhou, the study employs a deep convolutional neural network (CNN) for emotion analysis. This deep learning model is designed to automatically recognize emotional patterns indicative of MCI. The proposed method of detecting emotions and their persistence at various points of a watched video is significant in its potential to offer an automated and efficient tool for early detection of cognitive decline in the elderly. By leveraging the capabilities of deep learning, the authors aim to enhance the accuracy and reliability of MCI diagnosis. Emotion analysis serves as a unique marker for cognitive health, and the CNN is trained to discern subtle emotional cues that may be indicative of underlying cognitive impairments. This research aligns with the growing importance of leveraging advanced technologies to address health challenges associated with aging populations, ultimately contributing to the development of more effective and timely interventions for individuals at risk of MCI.

Conclusion:

Emotional AI is a growing field that can be leveraged to ease out adversarial situations, such as costly MRIs, sleepy drivers, closeted mental health, or the security of corporate secrets. I have particularly focused on surveillance and cognitive impairments, an odd mix (but not to Shanghai!). I have used this study to envision a system where a surveillance system could be used to help detect a cognitively impaired person who is lost. Consented surveillance is something that Big Tech companies are already pursuing discreetly, so why not use it to help families to detect an oft-elusive cognitive disease? Through these papers, I have studied the importance of context and a diverse dataset for better recognizing an expression, and have also studied how Shanghai's mental health centers use FER techniques to zero in on a cognitive impairment.

References:

1. 'Affective' computing and emotion recognition systems | Proceedings of the 2nd annual conference on Information security curriculum development
2. http://openaccess.thecvf.com/content_cvpr_2017/html/Kosti_Emotion_Recognition_in_CVPR_2017_paper.html
3. https://www.nature.com/articles/s41598-023-35446-4
4. https://www.sciencedirect.com/science/article/abs/pii/S0925231220300783