Real-time Driver's Emotion Recognition based on Deep Learning

Chao Zhang (cz2572)   Longwei Guo (lg3095)

Topic: A1: Face Recognition, Feeling Recognition, and Interaction

Date: 2020/2/14

Abstract: Studies have shown that driver's emotions play an important role in driving behavior. Negative emotions like anger, sadness and fear can have a significant impact on driver's response time which may cause fatal car accidents. Therefore, an emotion monitor and a timely warning to the driver will help maintain safety on the roads. Compared with the methods by using physical data from sensors, facial expression is a more straightforward feature related to emotions and also easier to monitor and capture. In our project, we propose a novel real-time emotion recognition system based on facial expressions to analyze a driver's emotion so as to ensure driving safety and reduce the risk of accidents affected by emotions. For this purpose, first we will design a data preprocessing module to extract face parts from background which helps reduce data complexity. Next we will architect several deep neural network structures to classify emotion into seven basic classes. Based on the optimal one, we will build an interface to detect driver's emotions in real time with webcam or mobile cameras. We aim to construct an efficient model that can recognize driver's expressions accurately and quickly. Our proposed method will be evaluated experimentally on different datasets including CK+, FER2013 as well as the real driving setting data KMU-FED and compared with other state-of-the-art methods.

1. Background (Review of Related Literature):

Emotions can influence human behavior a lot. At present, emotion recognition has been applied in various fields such as marketing, gaming and Human Machine Interface (HMI). Emotion detection of drivers is also a highly valuable task. Studies have established that the risk of an accident affected by negative emotions including anger and sadness is still high[1]. Therefore, emotion recognition can be used to give a timely alarm to the drivers to improve the safety of roads.

In the past few years, various techniques have been developed for monitoring a driver's emotional state. Some of them are based on physiological signals such as EEG, ECG, electrodermal activity, electromyography, etc[2]. To monitor these signals, it requires drivers to wear external devices or sensors have to be connected to the drivers' bodies, which may even cause distraction and disturbance during driving. However, cameras do not need drivers' active participation so that facial expressions can be captured for emotion recognition.

There are mainly two kinds of methods of which one is based on the hand-crafted features and the other is based on deep neural networks (DNNs). The hand-crafted features include Action Unit (AU), local binary pattern (LBP), a histogram of oriented gradients (HOG)[3], scale invariant feature transform (SIFT)[4], etc. Spiros et al.[5] proposed a framework based on facial animation parameters (FAP) to classify emotions. Chang and Chen[6] recognized FE using different AUs to describe face muscle movement. Gao et al.[7] proposed a real-time framework for driver's emotion recognition by extracting SIFT feature and using SVM for classification.

DNN based FER methods have various versions such as convolutional neural networks (CNN), long-short term memory (LSTM), generative adversarial networks (GANs)[8], inception and ResNet modules[9]. Since CNN based models usually learn quickly, it's very suitable for a real-time system. Hong-Wei et al.[10] use transfer learning in CNN to classify facial expressions. Hasani et al.[9] proposed a 3D convolutional networks model which contains 3D Inception-ResNet layers followed by LSTM layers to extract spatial and temporal features of facial expressions and frame relations in the videos. Mollahosseini et al.[11] proposed a DNN with 2 convolutional layers followed by max pooling layers and then four inception layers. This method gained a high accuracy of 93.3% on CK+ and 77.6% on MMI.

Based on the research we have done, for our project it's suitable to use CNN based neural networks to extract features and learn quickly which helps to ensure the quality of a real-time system.

2. Introduction to the Project:

2.1 Goal

In this project, we aim to develop a real-time system using deep neural networks that is able to analyze driver's emotion from facial expressions accurately and quickly. To generalize our model, we will test proposed models on both regular facial expression dataset and specific faces which are collected in an actual driving environment. There are two main functions in our system. First it can extract face parts and landmarks from image backgrounds of the video stream so that noise data can be removed as much as possible. Then based on a well-trained neural network model, it can classify the facial expression into seven basic emotion classes in real time.

2.2 Data preprocessing

There are several databases that we will use for our project. However, each database has different features and resolutions of data. For instance, the FER-2013 database has resized all the images into resolution pixels of 48×48 while CK+ database has pixel resolutions of 640×480 and 640×490. Since the FER-2013 database has been well preprocessed (all the face parts have been adjusted to the central region), we will focus on dealing with data from the CK+ and the KMU-FED database[12].

First, we will converse all the images into the grayscale values. Then we will use the Dlib toolkit and openCV method to detect human faces from image data and resize them into the same size. Then we can import pretrained models to attain facial features and landmarks for next steps. We will extract main face parts from backgrounds to reduce noise data.

And also, it's obvious that in the real driving environment the images cannot be as clear as those which we capture indoors. Considering such situations, we will also do some data augmentation to generate some data that contain different lightness, various face locations and even fuzzy images.

2.3 Model Architecture Design and Evaluation

Deep learning has popularity in application of image classification especially convolutional neural networks. Therefore, in our project we will focus on constructing several neural network models and use transfer learning to classify face expressions using Keras/Tensorflow.

In this part, we will use several pre-trained models with weights such as VGG16, ResNet50V2 and InceptionV3 and self-constructed CNN-based models to classify images and evaluate their performance. Through tuning parameters and architecture adjustment, we aim to get a relatively accurate model through a reasonable training process.

In the evaluation phase, we will firstly test our model on basic facial expression database FER-2013 and CK+. According to the results, we can analyze the generalization ability of our model. For the specific application background of our project, that is, driver emotion recognition, we will also test our model on KMU-FED data and some videos data from movie clips which contain driving settings. Although we cannot gather labeled movie videos, it's easy for us to judge the drivers' facial expressions ourselves and then analyze its accuracy of the predictions of the proposed model. We even do some comparison between our model and other state-of-art methods.

At last, we will use the webcams by calling Python methods to capture a real-time video stream. We can instruct several participants to perform different expressions and analyze the response time and accuracy of this real-time system.

2.4 Challenge

There are still many challenges in building our models.

First, for NN models it's prone to overfit when we do not use suitable training data and validation data or reasonable epochs of training. There is one way to address this is to use a huge amount of training data and try to make this data balanced and unbiased. And also, if the model structure is too complex, it's also likely to overfit with a small training set.

Also, the actual driving environments are very complicated. The quality of the images captured by the camera of the cars heavily depends on the location of the camera, the light of time, driving speed or even driver's actions and gestures. For instance, a driver's hands may cover his face and the light of noon may cause highly enhanced image contrast. This kind of data is very hard to be gathered and we can never cover all the possible situations that may occur on the roads. One way we can reduce such impact is to use high-quality cameras which can adjust lights and handle shake automatically.

What's more, considering it's a real-time system, the response time is a very important attribute that we should focus on. To reduce the time of processing image data, our model cannot be too large or too slow. Compared with LSTM, CNN is much faster. When constructing our own model, we will take it into consideration.

3. Introduction to the Dataset:

3.1 CK+ Database

The Extended Cohn-Kanade (CK+) database was released for the purpose of promoting research into automatically detecting individual facial expressions. It contains 327 sequences from 118 subjects which are labeled with seven discrete emotions ("Happy", "Surprise", "Sad", "Disgust", "Fear", "Anger" and "Contempt"). The participants photographed are at the age between 18 and 50 among which 31% are men and 69% are women. The sequences start from the neutral state to the apex state. All the sequence images also include facial landmarks and FACS code. These images have pixel resolutions of 640×480 and 640×490 with 8-bit grayscale values.

3.2 FER-2013

Facial Expression Recognition (FER) 2013 database was introduced during the ICML 2013 Challenges in Representation Learning. It was collected automatically by using Google image search API. It contains 35,887 grayscale face images with pixel resolutions of 48×48. For a

Kaggle competition, it is further divided into a training set containing 28,709 images and a test set containing 3,589 images. All the faces have been registered and adjusted to the cropped region. These images are categorized into seven classes of emotions among which 4,593 images are labeled with "angry", 547 images with "disgust", 5,121 images with "fear", 8,989 images with "happy", 6,077 images with "sad", 4,022 images with "surprise" and 6,198 with "neutral". Although the data has been preprocessed well, because of its low resolution, it's difficult to extract facial landmarks accurately.

3.3 KMU-FED

To improve the effectiveness of our proposed method for driver emotion recognition, we also need a lot of data which are collected in a real driving environment. KMU-FED database is for FER in an actual driving set even including some problems that may occur on a real-life road. The sequences in this dataset were captured in a real driving vehicle environment with an NIR camera which was installed on the dashboard or steering wheel in the car. It consists of 55 image sequences from 12 subjects which include various changes in illumination and partial unclearness caused by hair of glasses. The images have pixel resolutions of 1600×1200.

3.4 Other data

Although we choose a dataset collected in a driving environment, the data is still limited. We also need to test our methods using driving data. To address this problem, we will collect some data by ourselves. Since CK+ data is collected in a laboratory setting, the expressions cannot be as spontaneous as those in usual daily life. And also, there are many factors such as the locations of the camera, the various lightness of time and different gestures of drivers affecting the accuracy of emotion recognition. Therefore, we enlarge our dataset through gathering movie clips and videos on the internet ourselves which contain more continuous actions of drivers and more comprehensive situations that may happen on the roads.

4. Plan:

4.1 Milestone 1 (Feb.14th)

- Jan.28th - Feb. 7th:
  Literature review and past related work summary
- Feb.8th - Feb.14th:
  Project organization and proposal writing;

4.2 Milestone 2 (Mar.13rd)

- Feb.15th - Feb.29th:

Dataset collection and downloading; Image data preprocessing (grayscale conversion, size normalization and data augmentation); Environment configuration (Python, OpenCV, Keras and Dlib…)

- Mar.1st - Mar.13rd: Face and facial landmark extraction; Neural network architecture design and training; Presentation and report writing

4.3 Milestone 3 (Apr.17th)

- Mar.14th - Mar.30th:
  Model improvement and parameter tuning; Cross dataset evaluation and test
- Apr.1st - Apr.17th:
  Real-time interface design; Presentation and report writing

4.4 Final (May 8th)

- Apr.18th - May 8th:
  Real setting video test; demo and final report writing; Presentation and report writing

Reference:

[1] Kowalczuk, Z., Czubenko, M. and Merta, T. (2019). Emotion monitoring system for drivers. IFAC-PapersOnLine, [online] 52(8), pp.200–205.

[2] C.D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, Toward emotion recognition in car-racing drivers: A biosignal processing ap- proach, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 38, no. 3, pp. 502–512, 2008. Orrite, C.; Ganan, A.; Rogez, G. Hog-based decision tree for facial expression classification.

[3] Greche, L.; Es-Sbai, N. Automatic System for Facial Expression Recognition Based Histogram of Oriented Gradient and Normalized Cross Correlation. In Proceedings of the 2016 International Conference on Information Technology for Organizations Development (IT4OD), Fez, Morocco, 30 March-1 April 2016.

[4] Barroso, E.; Santos, G.; Proenca, H. Facial expressions: Discriminability of facial regions and relationship to biometrics recognition. In Proceedings of the 2013 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), Nanyang, Singapore, 16-19 April 2013; pp. 77-80.

[5] S.V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias, Emotion recognition through facial expression analysis based on a neurofuzzy network, Neural Networks, Vol. 18, no. 4, pp. 423–435, 2005.

[6] Chang, J.Y.; Chen, J.L. A facial expression recognition system using neural networks. In Proceedings of the International Joint Conference on Neural Networks, Washington, DC, USA, 10-16 July 1999; pp. 3511-3516.

[7] H.Gao, A. Yu ̈ce, and J.P. Thiran, Detecting emotional stress from facial expressions for driving safety,Image Processing (ICIP), 2014 IEEE International Conference on, pp. 5961–5965, 2014.

[8] Yang, H.; Ciftci, U.; Yin, L. Facial Expression Recognition by De-expression residue Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18-22 June 2018; pp. 2168-2177.

[9] Hasani, B.; Mahoor, M.H. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21-26 July 2017; pp. 2278-2288.

[10] H.W. Ng, V.D. Nguyen, V.Vonikakis, and S.Winkler, Deep learning for emotion recognition on small datasets using transfer learning, Pro- ceedings of the 2015 ACM on international conference on multimodal interaction, pp. 443–449, 2015. A.

[11] Mollahosseini, D. Chan, and M.H. Mahoor, Going deeper in facial expression recognition using deep neural networks, Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp. 1– 10, 2016.

[12] Jeong M, Ko BC. Driver's Facial Expression Recognition in Real-Time for Safe Driving. Sensors (Basel). 2018;18(12):4270. Published 2018 Dec 4. doi:10.3390/s18124270