Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach

Jinso Raj

Govt. Engineering College Barton Hill, Thiruvananthapuram.

Dec 7 2021

Contents

Introduction

- DRIVER is in the center of the Road-Vehicle-Driver loop.
- Driver decision and behaviors are the major aspects that can affect driving safety.
- The accident rate can be reduced by 10% to 20% with a precise driver behavior monitoring system.
- Therefore, the recognition of driver behaviors is becoming one of the most important tasks for intelligent vehicles.

- Regarding the intelligent and highly automated vehicles, such as the Level-3 automated vehicles, the driver is responsible for taking over the vehicle control under emergencies.
- At this moment, the real-time driver behavior and activity monitoring system has to decide whether the driver can take over or not.
- The recognition models are trained to identify seven common driving-related tasks and also to determine whether the driver is being distracted or not.
- With this end-to-end approach, intelligent vehicles can better interact with human drivers and properly making decisions and generating human-like driving strategies.

Related work

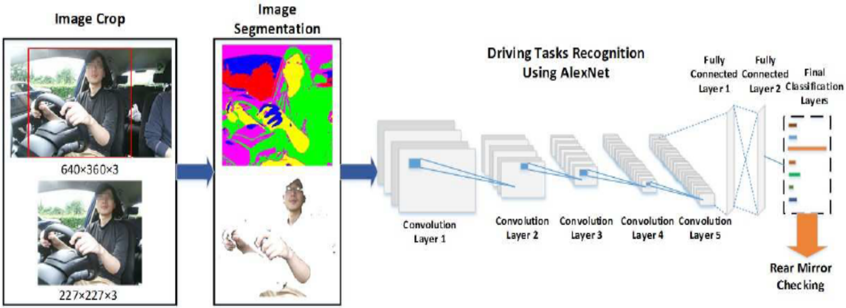
- Driver behaviors have been widely studied over the past two decades.
- Previous studies mainly focus on the driver attention and distraction (either physical distraction or cognitive), driver intention , driver styles,driver drowsiness and fatigue detection.
- The physiological signals, such as the electroencephalogram (EEG) and electrooculography (EOG) are also widely used for real-time driver status monitoring.
- Most of the existing driver behavior studies require extracting specific features in advance These features are not always easy to be obtained, and some even require specific hardware devices, which will increase either the temporal or the financial cost.

- In this work, three different CNN models will be evaluated for driver activities recognition and distraction detection tasks.
- The only sensors required in this study is low cost RGB camera.
- The CNN models take the processed images directly without any manual feature extraction procedure. By applying the transfer learning scheme, the pre-trained CNN models can be efficiently fine-tuned to satisfy the behaviors detection task.

System overview

- This system can be summarized as follows.
    1. A deep learning-based approach is applied to identify driver behaviors. Unlike existing studies that require complex algorithms to estimate the driver status information, the proposed algorithm takes merely the color images as the input and directly outputs the driver behavior information.
    2. Transfer learning is applied to fine-tune the pretrained deep CNN models. models are trained to deal with both the multiple classification tasks and the binary classification task.
    3. An unsupervised GMM-based segmentation method is applied to process the raw images and extract the driver body region from the background.

Design of the system



Design

- Raw RGB images are collected using the Kinect camera.
- The cropped images are segmented using the GMM algorithm.
- Finally, the CNN model is adopted for the activities recognition task.

System setup

Fig. 2. Experiment setup. The Kinect is mounted on the middle of the front window and data are collected using a laptop.

- The Kinect enables the collection of multi-modal signals, such as the color image, depth images.
- The sampling rate for the image collection is 25 frames per seconds.
- The data are recorded with an Intel Core i7 2.5GHz CPU.
- To store the images, the raw images are compressed to 640*360*3 format to increase the computation efficiency.

Algorithms used in the system

- **GMM algorithm**: it is used for the image preprocessing and the segmentation.
- Three deep CNN framework used for model preparation.
  1. AlexNet Model
  2. GoogLeNet Model
  3. ResNet Model

GMM

- The GMM algorithm is applied to segment the images and extract the driver body region from the background.
- GMM is a probability density function that is represented by a weighted sum of sub-Gaussian components.
- To train a GMM-based segmentation model, each image is represented by a feature vector according to the pixel intensity.
- The feature vector for the GMM is a three-dimensional vector that contains the RGB intensity of each pixel.

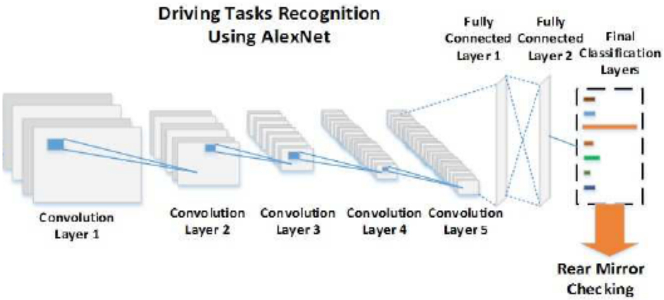
- A Gaussian Mixture is a function that is comprised of several Gaussians, each identified by $k \in \{1, \ldots, K\}$, where K is the number of clusters of our dataset.
- GMM will find probability for each class for a data point.
- In other word we can say that, GMM is a mixed representation of diferent Guassians.
- The algorithm consist of 2 primary steps, Expectation(E) and Maximization(M).
    1. Expectation step - For each data point, probability for each guassian component is calculated.
    2. Maximization step - Adjust the Mean , Variance and weight of each data point to maximize the possibility of getting these parameters.

Image Pre-processing and Segmentation

Illustration of the raw images and segmented images.

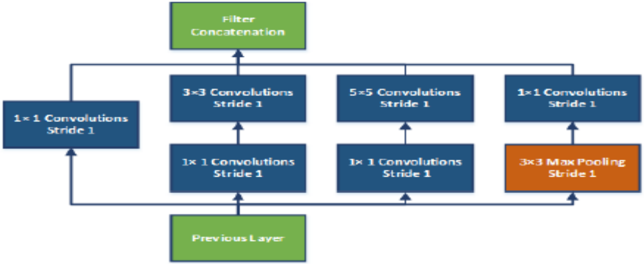
AlexNet model:

- ImageNet is a large-scale dataset, which contains more than 15 million high-resolution annotated natural images of over 22,000 categories.
- There are five convolutional layers and three fully connected neural network layers with non-linearity and pooling layers between the convolutional layers.
- AlexNet contains 60 million parameter and 650,000 neurons



GoogleNet Model:

- GoogleNet is significantly deeper than the AlexNet, and it achieved more accurate classification results on the ImageNet dataset.
- The main contribution of GoogLeNet is the utilization of Inception architecture.
- Each Inception layer consists of six basic convolution filters and one max pooling filter.



ResNet Model:

- ResNet enables the construction of deeper convulation neural network by introducing the residual learning scheme.
- By introducing the identity mapping and copying the other layers from the shallower model, the deep residual network can efficiently solve the model degradation problem when the models getting deeper.

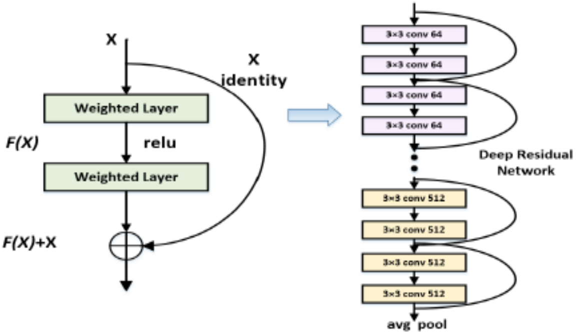


TABLE I
CLASSIFICATION RESULTS FOR DRIVING TASKS RECOGNITION USING ALEXNET

No.	GMM Based AlexNet							
	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.825	0.929	0.011	0.225	0.840	1.0	0.972	0.771
D2	0.875	0.234	0.571	0.229	0.516	0.928	0.836	0.813
D3	0.564	0.684	0.0	0.711	0.747	0.983	0.983	0.908
D4	0.825	0.469	0.927	0.399	0.0	0.958	0.994	0.786
D5	0.797	0.20	0.10	0.843	0.60	0.959	0.996	0.843
D6	0.957	0.928	0.852	0.977	0.783	0.926	0.999	0.928
D7	0.993	0.921	0.915	0.951	0.913	0.290	0.981	0.878
D8	0.990	0.989	0.417	1.0	0.991	0.996	0.736	0.880
D9	0.353	0.994	0.229	0.813	1.0	0.982	0.979	0.752
D10	0.528	0.724	0.447	0.798	0.274	1.0	0.995	0.684
Mean	0.786	0.869	0.545	0.802	0.771	0.932	0.945	**0.816**

TABLE II
CLASSIFICATION RESULTS FOR DRIVING TASKS RECOGNITION USING GOOGLENET

No.	GMM Based GoogLeNet							
	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.917	0.619	0.0	0.325	0.433	1.0	0.968	0.768
D2	0.892	0.362	0.0	0.042	0.230	0.784	0.815	0.767
D3	0.883	0.563	0.0	0.073	0.840	1.0	0.994	0.739
D4	0.740	0.453	0.848	0.986	0.758	0.663	1.0	0.755
D5	0.970	0.20	0.233	0.325	0.078	0.959	0.988	0.799
D6	0.951	0.966	0.807	0.936	0.967	0.075	1.0	0.829
D7	1.0	0.886	0.436	0.990	0.890	0.248	0.963	0.737
D8	0.301	0.995	0.178	1.0	1.0	0.990	0.998	0.789
D9	0.562	0.245	0.949	0.997	1.0	0.990	0.843	0.792
D10	0.990	1.0	1.0	0.685	0.882	0.012	1.0	0.810
Mean	0.835	0.766	0.648	0.796	0.819	0.678	0.948	**0.786**

TABLE III
CLASSIFICATION RESULTS FOR DRIVING TASKS RECOGNITION USING RESNET50

No.	GMM Based ResNet50							
	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.944	0.389	0.120	0.125	0.219	1.0	0.963	0.746
D2	0.872	0.284	0.0	0.729	0.066	0.918	0.926	0.921
D3	0.919	0.938	0.195	0.040	0.814	0.998	0.993	0.753
D4	0.975	1.0	0.924	0.514	1.0	0.639	0.882	0.801
D5	0.907	0.255	0.133	0.874	0.473	0.930	0.996	0.856
D6	0.790	0.992	0.941	0.791	0.504	0.509	0.985	0.750
D7	0.996	0.857	0.629	0.922	0.950	0.301	0.973	0.786
D8	0.528	0.567	0.192	0.641	0.988	0.944	0.715	0.638
D9	0.346	0.245	0.713	0.997	0.735	0.693	0.829	0.655
D10	0.002	0.999	0.058	0.991	0.782	0.219	1.0	0.589
Mean	0.728	0.652	0.391	0.662	0.653	0.715	0.926	**0.749**

Advantages of the system

- It is reported that more than 90% light vehicle accidents are caused by human driver misbehavior in the United States, and the accident rate can be reduced by 10% to 20% with a precise driver behavior monitoring system.
- Real-time Application:the system is implemented to a Windows operating system using MATLAB platform and a single lowcost GPU device.
- The testing cost of the AlexNet for each image is about 13ms, also, it cost 50ms for the GMM to segment each image. The total computational cost for each image is around 60-70ms, and the general processing ability of the system is about 14fps.

Conclusion

- In this system, a driving-related activity recognition system based on the deep CNN model and transfer learning method is proposed.
- To increase the identification accuracy, the raw RGB images are first processed with a GMM-based segmentation algorithm, which can efficiently remove the irrelevant objects and identify the driver position from the background context.
- The classification results indicate that the segmentation contributes to a much more precise detection result than the model trained with the raw images.

Future work

- The data will be further analyzed, and the model will be updated to increase the system robustness and detection accuracy.
- Meanwhile, the system will be tested and used for driver/passenger behavior analysis on the partially automated vehicles in the real world.

References

- Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach - https://ieeexplore.ieee.org/document/8678436/
- Color Image Segmentation Using Gaussian Mixture Model and EM Algorithm - link.springer.com/chapter/10.1007/978-3-642-35286-7_9
- Gaussian Mixture Model - https://brilliant.org/wiki/gaussian-mixture-model/

Thank You