Comparative Study of Dashcam-Based Vehicle Incident Detection Techniques

Farzanuddin Hisam Ahmad
University of Houston
Houston, USA
e-mail: fhahmad@uh.edu

Yunpeng Zhang
University of Houston Houston, USA
e-mail: yzhan226@central.uh.edu

Fengxiang Qiao
Texas Southern University
Houston, U.S.A.
e-mail: fengxiang.qiao@tsu.edu

***Abstract*— With the increase in world population demand for the vehicle is also growing which is thus leading to a higher number of crashes on road. These crashes lead to personal, economical and property losses. To make roads safer and reduce the number of crashes on the roads, intense studies are being carried out to anticipate, prevent, and detect incidents. These researches use data collected from video cameras and sensors installed on roadsides and within vehicles. The crash data obtained from these sources can be used to investigate the severity of the situation and human behaviors. In this paper, various video-based road incident identification techniques are overviewed that used dashcams installed within vehicles as their data source. This paper puts forward an outline of the current works in the area of vehicular crash detection and underlines the drawbacks and prospects of the study.**

I. INTRODUCTION

As per the National Security Council, 4.4 million people were injured and 38,800 deaths were recorded in vehicle accidents in 2019 in the U.S. alone [1]. The vehicular crashes come with accidental inquiries, lawsuits and insurance delays in addition to injuries and personal losses. Advancement in technology and digitalization has made the acquisition of such huge data of road incidents easy. Analyzing and studying the trends of these data has also become less time-consuming. These data are like digital resources that can be utilized to develop better transportation systems and smart infrastructures that will reduce the incident risks on the roads. Surveillance videos obtained from traffic cameras and dashboard cameras which are based on vision-based technology are the most commonly used ones to gather information related to transportation.

Additionally, to give a better sense of safety, nowadays most vehicle manufacturers have equipped their vehicles with dashboard cameras (dashcams). Moreover, the owners of old vehicles are also spending money on dashcams. This widespread installation and usage of dashcams have now given us access to a pool of video data. These videos contain recordings of vehicular movement and human behavior in different ways.

Out of the numerous reasons, distracted driving is the leading cause of road incidents [2]. These distractions might be because of talking to fellow passengers, answering phone calls, replying to messages etc. Forecasting and notifying drivers about future dangerous events like incidents can significantly reduce the number of vehicular crashes. There are numerous ways to tackle this issue. Some of the common approaches used for this purpose utilize data retrieved from closed-circuit television (CCTV). cameras, magnetic sensors, proximity sensors, etc. CCTV cameras are generally installed on traffic lights, buildings, walls, electric poles etc.

Since dashcam video data are easy to access and visualize, intensive studies are being carried out to develop incident prediction and prevention systems utilizing this form of data. The objective of this paper is to evaluate various vision-based incident prediction and detection techniques for their versatility and effectiveness. An overview of existing works in the field is provided with an understanding of their drawbacks and prospects.

The remainder of the paper is ordered as follows. Section II defines various approaches to incident detection based on dashcam data. Section III compares the results from each of the studies. Section IV concludes the paper.

II. INCIDENT DETECTION USING DASHCAM VIDEOS

In the past few decades, numerous crash detection techniques have been put forward by scientists and researchers. The techniques proposed earlier used data from different sources like motion sensors of the car, navigation systems, street cameras, etc. Road incidences detection can be carried out using data from each of these sources. However, in this paper, we have limited our survey to the studies that have been carried out using dashcam videos as their primary data source for proposing a crash detection system. The reason for including only dashcam video-based techniques is that dashcam videos are easily available on the open platform to be downloaded and can also be easily be visualized by anyone.

The existing crash detection techniques from kinds of literature can be summarized into the following categories:

a) Visual Information Tracking,
b) Event Detection and classification.
c) Dynamic Spatial Attention Recurrent Neural Network
d) Object's future location prediction.

To analyze the techniques mentioned above, one paper based on each technique is chosen. The criteria for selecting these papers were based on their effective usage of dashcam videos for detecting road incidences.

These techniques are explained, analyzed and concluded in the following subsections.

A. Visual Information Tracking

Various approaches have been developed in the existing literature to detect automobile crashes on roads. These approaches can be broadly classified as vision-based and non-vision-based methods. In the non-vision-based method, data is derived from specialized sensors installed in the vehicles. However, the non-vision based approaches proposed in [3], [4] and [5] may return false positives due to the complex and interdependent nature of the real world.

To eliminate the chances of getting false negatives, the vision-based approach is implemented which uses specialized types of equipment like mounted cameras. A vision-based approach uses both audios as well as videos to identify different traffic situations. Dashboard video cameras are readily available and effortless to mount.

1) Methodology

An immediate change in the visual information is used in [6] to detect accidents using the dashcam video *(V)*. The difference in frames (t) of *V*, where *t* is the time, is calculated by equation (1). These differences include the changes in Red (R), Green (G) and Blue (B) components of the successive frames.

$$\mathrm{df(t)} = |\mathrm{V(t+3)} - \mathrm{V(t)}| \tag{1}$$

$$\Pi = (\mathrm{F}, \mathrm{V}, \Omega) \tag{2}$$

In equation (2), (Ω) and (*F*) are threshold omega and algorithm for accident detection, respectively, which are used to characterize the accident detection model (*Π*).

Then, a range of grey values is obtained by constructing the histogram of frame difference calculation for all three-color planes of the colored videos. Pixels with a drastic change in their intensities are counted by aggregating frequency information in the range of 100 to 200. This aggregation (t) will be high during an accident, and its occurrence is determined by (Ω).

The incident detection algorithm selects every 3 frames and calculates the difference between each successively selected frame. The dashboard video accident detection (DVAD) is used to perform an analysis of 60 videos, each representing different accident scenarios

2) Analysis

A dataset was compiled from Youtube videos involving road incidences recorded in dashcams. This approach was able to recognize 38 accidents video correctly out of 40 videos, and all the 20 non-accidental videos were also correctly recognized. These results were made when the value of Ω was set to be $2.5\mathrm{x}10^5$. The efficiency of the approach is dependent on the optimal Ω value with maximum efficiency with the threshold value in the range of 190-250.

The variation of the accuracy of this model with respect to changes in the threshold values is shown in Fig. 1.

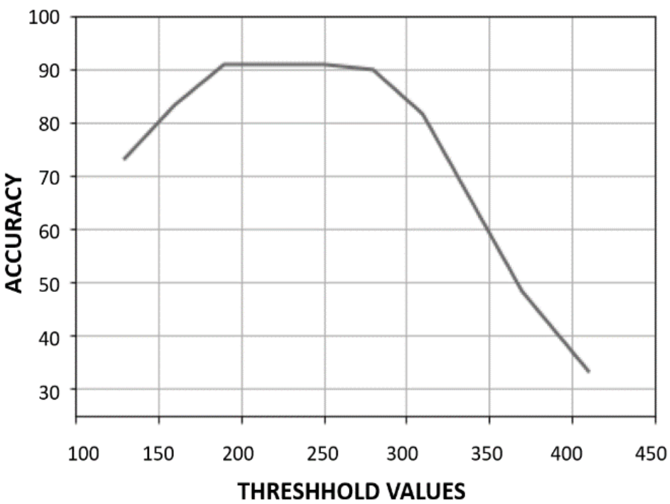

Fig. 1. Threshold values vs accuracy graph [6]

3) Conclusion

This vehicle crash detection approach is based on dashcam videos with a computer vision technique by monitoring changes in visual information. The crashes which happened in fractions of a second can also be detected using this approach. The overall efficiency of 96.6% was recorded for this approach.

B. Event detection and classification

Detection of treacherous events such as incidents and near-incident scenarios on streets and highways is a primary task in developing smart and safe transportation system for road users. To detect and categorize such events, [7] puts forward a technique that combines machine learning techniques with computer vision. This method is applied to dashcam videos and the telematics data SHRP2 [8] datasheet is used for the cataloging process.

1) Methodology

Random forest classifier [9] is used to merge features from several sources and catalog events after extraction of high-leveled features. The Convolution Neural Network model [9] is used for every event to recognize the objects in every video frame. The feature extractor is fed with intermediate features after running a pipeline to calculate their transitional features. The optical flow of vectors is used to obtain other features. Ground truth labels resulting from the SHRP dataset utilizing these features are used to train the random forest model [7].

YOLOv3's [10] modified version is trained on the dataset of COCO [11], is used to forecast bounding boxes for objects belonging to the classes of cars, bicycles, persons, busses, trucks, motorbikes, stop sign and traffic light. Each convolution block comprises of: (i) convolution filter, (ii) nonlinear activation function, and (iii) max-pooling layer. The forecast is conducted at 3 distinct scales, and objects are categorized at distinct scales.

The vanishing point is approximated, and the collation cone is created with the dense optic flow, which is computed by keeping track of displacement vectors from frames of t-1 to t. Farnebäck Algorithm [12] was used on OpenCV. Afterward, Time-To-Contact (TTC) of objects is obtained by processing the data which is used to establish their chance of crash with the subject vehicle. YOLO was used to perform object tracking. TTC was calculated using equation (3) after applying low pass filter for each tracked object.

$$TTC = \frac{\Delta t}{\left(\frac{\omega - \omega'}{\omega'}\right)} \tag{3}$$

Where Δt is the difference in time, ω *and* ω' are box width in frames and $\left(\frac{\omega-\omega'}{\omega'}\right)$ represents scale change of the bounding box which may also be represented as s.

The vanishing point was estimated by using a procedure based on RANSAC then the collision cone was computed. The tracked object is flagged if it is predicted to enters the collision cone of the ego vehicle.

The features acquired from dense optic flow and road scene pipeline are computed for the final classifier. Almost 200 features are calculated which are then utilized to train Random Forest Classifier.

Figure 2 shows a simplified flow chart of the proposed crash detection and classification procedure. Video data from dashcam is fed into Dense Optic Flow (DOF) and YOLOv3 separately. Then the output from both of them is combined and sent to Road Scene Pipeline (RSP). Then, Telematic Data is incorporated with the processed data received from RSP, and various features are extracted for easy classification and prediction, which is done by RFC.

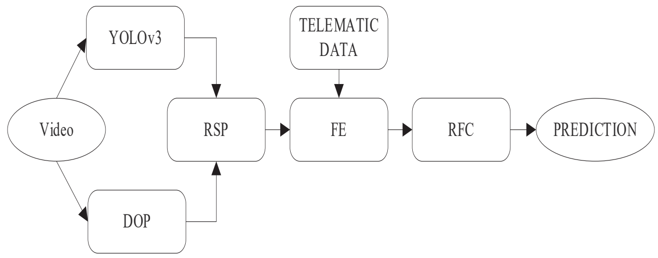

Fig. 2. Classification flow diagram [7]

2) Analysis

To evaluate the performance of this approach, two types of categorization analysis was conducted. The first one was to create a two-class problem that will use two-level groups in which the event can be classified into *i.e.*, crash events and near-crash events. Similarly, the second one was to create a three-class problem which will have one extra category *i.e.*, crash, near-crash and safe events.

The global efficiency of 87% was achieved for the 2-class problem and 85% average accuracy per class. Similarly, in the 3-class problem, the global accuracy was more than 85%. However, the average accuracies per class were not close to each other.

3) Conclusion

Using videos from dashcams and telematics data a new approach to identifying and classifying vehicular road incidents was developed. Object detection, machine learning and computer vision techniques were combined to build this system. Two-category and multi-category experiments were carried out with an efficiency of 87% and 85%, respectively.

C. Dynamic Spatial Attention Recurrent Neural Network (DSN-RNN)

Predicting incidents beforehand is one of the most crucial steps in developing a safer transportation system. To achieve this, [13] introduced a Dynamic-Spatial-Attention (DSA) Recurrent Neural Network (RNN) which is used the Long Short-Term Memory (LSTM) cells. A unique dataset consisting of incidences involving bike hits, car hits etc., was downloaded and used for the analysis of the proposed model.

1) Methodology

RNN accepts observational sequences as input and gives learned hidden representational sequences as output. Using softmax function probability of events is calculated. There is a vanishing gradient issue when working with RNN, therefore LSTM cells are used to avoid this problem.

Vehicles, pedestrians, and other similar objects were the observations from videos, on which the RNN concentrated to predict the incidents. To calculate the spatially specific objects' dynamic weighted, the soft attention mechanism [14] was used. To calculate the relevance amongst the previously unseen representation and every spatial-specific observation, unnormalized attention weights were designed using equation (4):

$$e_t^j = w^T \rho(W_e \mathrm{h}_{t-1} + \mathrm{U}_e \widehat{x_t^j} + \mathrm{b}_e \quad (4)$$

Then full frame features and spatially specific object features are combined. The sum of entropy losses and exponential losses was used to describe the loss function. This method was used to train and detect incidents on TensorFlow [15] in the dashcam videos. The probability of future accidents is generated to analyze the accuracy of the method.

2) *Analysis*

Without modeling frames relation, the following variant of methods using baseline ones and RNN was compared.

- Dynamic Spatial Attention RNN
- Average Attention RNN
- Average Attention Single-frame Classifier
- Max Probability Single-frame Classifier
- Frame base Single-frame Classifier

Snapshots of some of the videos from the dataset on which the model was analyzed is depicted in following Fig 3 and Fig 4. The model was successfully able to identify the road scenarios.

Fig 3. Car hitting bike [13]

Fig 4. Car hitting another car [13]

To compare the effectiveness of the VGG appearance and IDT motion feature, all the methods were separately analyzed by both VGG and IDT. Then, the best variants from both were extracted and fused.

3) *Conclusion*

The crash detection technique using the Dynamic Spatial Attention RNN model is claimed to outperform other previously developed baselines without RNN. This approach can predict an incident 2 seconds before its actual occurrence. Precision and recall values were recorded as 56.14% and 80%, respectively.

D. Object's future location prediction

Comparing the two consecutive frames and identifying the differences in a video can be used to detect accidents [16]. However, the anomalies detected may not always imply the occurrence of an incident in the frame, as the difference in the frames is due to the change in the position of the object as well as the movement of the ego vehicle. To avoid this complication [17], only those changes are considered as anomaly whose observed trajectory deviates from the predicted trajectory.

The observed trajectory is extracted from the video frames by identifying the vehicles in the particular frame and then enclosing them in the box. The future location of the observed vehicle is calculated using Yolov3. This model is trained using diverse data set by following including daytime and nighttime driving scenarios. [16]. Additional input using ego-motion prediction [18] is also introduced which takes into consideration of changes in features of consecutive frames due to the motion of the vehicle on which the dashcam is installed.

The proposed unsupervised crash detection approach is streamlined into a flow chart (Fig 5). This approach uses three methods i.e.,

a) The predicted bounding box accuracy method
b) The predicted bounding mask accuracy method
c) The predicted bounding box consistency method

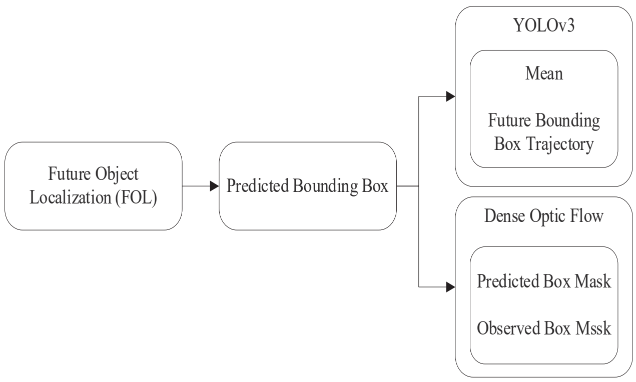

Fig 5. Unsupervised crash detection approach

1) *Methodology*

The current position of the observed object and its predicted trajectory is enclosed in separate boxes. It is based on a network structure described in [19]. This process is faster as it does not process information from previous frames rather uses information only from the current frame. The Region-of-Interest POOLing (RoIPool) operation is used to obtain spatiotemporal features.

It is important to collect information on ego-motion [18], [20] for predicting the future of the object accurately which is achieved by using RNN encoder-decoder. The FOL- Track Algorithm is used to overcome the false negatives and false alarms triggered due to occluded objects.

To detect crashes in the videos, three different strategies viz., predicted bounding boxes, predicted box masks and predicted bounding boxes were proposed by [17]. An Accident Detection (A3D) dataset was introduced and used for the evaluation of this method on PyTorch.

If the object traces the path predicted by the model then it is reported as a normal scenario. However, if there is a deviation of the object's original path from its predicted path then, the anomaly is reported by the model.

To detect an anomaly, the K-Nearest Neighbor Distance (K-NN), Conv-AE [13], and State-of-the-art [16] baselines were implemented for result comparisons.

2) Analysis

The AUC (area under the curve) method was to analyze the FOL (Future Object Localization) approach. It was found out that all the variations of the FOL outperformed the existing approach like KNN, Conv-AE and State-of-the-art [16]. Out of all the variations, the FOL-MaxSTD showed the best results. However, it may fail when it is difficult to predict future object localization.

The video is input into the FOL from where the future predicted location of the object is enclosed in the box. YOLOv3 is used to enclose the observed position of the object in the video. The predicted box is then matched with the observed position of the object. The maximum overlap of the box suggests the high accuracy of the model.

3) Conclusion

This crash detection method observes the current trajectories of the objects and use them to forecast the future position of the object. The predicted future location was checked against the observed future location. When an anomaly was detected, the scene was analyzed for the incidents. This approach is claimed to perform better than previously published baselines.

III. Result

Table I summarizes the key points of all the four video-based incident detection techniques mentioned in this paper. The efficiency of each of these techniques is compared with each other.

Additionally, remarks are provided that givers an overview of each model and describe their dependencies on factors like threshold value, class number etc. that highly influence the efficiency of the models.

IV. Conclusion

In this paper, we have successfully compared four techniques that can be used for road incidence detection and prediction using video clips from dashboard-mounted cameras. Each of them has its strengths and weaknesses.

TABLE I . EFFICIENCY COMPARISON

TECHNIQUES	EFFICIENCY	REMARKS
Track changes in visual information	96.6% For the omega value of 2.5 x 10^5	Threshold omega's (Ω) optimum value affects the efficiency
Event detection and classification	87% for 2-class problem	In a 2-class problem, the classifier achieves 0.94 recall.
	93% for 3- class problem	In a 3-class problem, the classifier achieves 0.49 recall.
Dynamic spatial attention recurrent neural network	74.35 average precision	Achieved by the combination of best IDT and VGG methods.
Predicting future location of objects	84.2% more accuracy than	Outperformed all baselines Showed false negatives in the cases of bike incidents

It was found that the model proposed by [6] can generate the most accurate results given that the value of the omega is carefully selected. The detection and classification of events by [7] can achieve high accuracy if the cataloging criteria are kept simple. The incidents are anticipated before their occurrence by [13] but with slightly low precision. The future location of the objects was well predicted in [17] and is claimed to outperform all the baselines. However, it misses the incongruities related to bike traffic.

V. Future work

All reviewed approaches tried to address the problem of vehicle crash detection using videos from dashcams. However, there is still room for improvement with even better performance.

The future studies can be concentrated on the following areas:

(1) Improve object detection on the roads and refine trajectory predictions.

(2) Increase the accuracy and precision of incident prediction.

(3) Incorporate more events in the classification for a better forecast.

(4) Reduce the time of computation and increase the time between the prediction and actual incident.

VI. Acknowledgment

This work is funded by Deep Learning-Based Intrusion Detection Approaches for Advanced Traffic Management Systems, Data Science Institute, University of Houston, U.S.A.

VII. References

[1] "Fatality Estimates." https://www.nsc.org/road-safety/safety-topics/fatality-estimates (accessed Aug. 17, 2020).

[2] andrew.currin.ctr@dot.gov, "Distracted Driving," *NHTSA*, Sep. 08, 2016. https://www.nhtsa.gov/risky-driving/distracted-driving (accessed Aug. 17, 2020).

[3] S. Tapadar, S. Ray, H. N. Saha, A. K. Saha, and R. Karlose, "Accident and alcohol detection in bluetooth enabled smart helmets for motorbikes," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 584–590.

[4] A. Shaik *et al.*, "Smart car: An IoT based accident detection system," in *2018 IEEE Global Conference on Internet of Things (GCIoT)*, 2018, pp. 1–5.

[5] T. H. Yee and P. Y. Lau, "Mobile vehicle crash detection system," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, Jan. 2018, pp. 1–4, doi: 10.1109/IWAIT.2018.8369671.

[6] E. P. Ijjina and S. K. Sharma, "Accident detection from dashboard camera video," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2019, pp. 1–4, doi: 10.1109/ICCCNT45670.2019.8944520.

[7] L. Taccari *et al.*, "Classification of Crash and Near-Crash Events from Dashcam Videos and Telematics," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2460–2465.

[8] J. M. Hankey, M. A. Perez, and J. A. McClafferty, "Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets," Apr. 2016, Accessed: Aug. 18, 2020. [Online]. Available: https://vtechworks.lib.vt.edu/handle/10919/70850.

[9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[10] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement. CoRR abs/1804.02767 (2018)," *ArXiv Prepr. ArXiv180402767*, 2018.

[11] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014, pp. 740–755.

[12] G. Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion," in *Image Analysis*, vol. 2749, J. Bigun and T. Gustavsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 363–370.

[13] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun, "Anticipating Accidents in Dashcam Videos," in *Computer Vision – ACCV 2016*, vol. 10114, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds. Cham: Springer International Publishing, 2017, pp. 136–153.

[14] K. Xu *et al.*, "Show, Attend and Tell: Neural Image CaptionGeneration with Visual Attention," p. 10.

[15] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *ArXiv160304467 Cs*, Mar. 2016, Accessed: Sep. 02, 2020. [Online]. Available: http://arxiv.org/abs/1603.04467.

[16] W. Liu, W. Luo, D. Lian, and S. Gao, "Future Frame Prediction for Anomaly Detection - A New Baseline," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 6536–6545, doi: 10.1109/CVPR.2018.00684.

[17] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins, "Unsupervised Traffic Accident Detection in First-Person Videos," *ArXiv190300618 Cs*, Jul. 2019, Accessed: Sep. 02, 2020. [Online]. Available: http://arxiv.org/abs/1903.00618.

[18] Y. Yao, M. Xu, C. Choi, D. J. Crandall, E. M. Atkins, and B. Dariush, "Egocentric Vision-based Future Vehicle Localization for Intelligent Driving Assistance Systems," *ArXiv180907408 Cs*, Mar. 2019, Accessed: Sep. 02, 2020. [Online]. Available: http://arxiv.org/abs/1809.07408.

[19] M. Xu, M. Gao, Y.-T. Chen, L. S. Davis, and D. J. Crandall, "Temporal Recurrent Networks for Online Action Detection," p. 10.

[20] A. Bhattacharyya, M. Fritz, and B. Schiele, "Long-Term On-board Prediction of People in Traffic Scenes Under Uncertainty," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, pp. 4194–4202, doi: 10.1109/CVPR.2018.00441.

[21] S. Zhendong, Y. Lei, and L. Lei, "A Method Study on the Target Value of Prevention of Truck Rear Under-Run Protection Device," in *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, Jan. 2017, pp. 144–147, doi: 10.1109/ICMTMA.2017.0043.