

Application of the Motion Detector and the Artificial Neural Network to Detect Vehicle Collisions: A Case Study

Artem Atoyan¹, Oleg Zolotov² and Yulia Romanovskaya^{1,*}

¹Murmansk State Technical University, Department of Math, Information Systems and Software Engineering, RU-183010, Murmansk, Russia

²Murmansk Arctic State University, Near-Earth Environment Computer Modelling Laboratory, RU-183025, Murmansk, Russia

Abstract. Motor vehicle collisions are a common cause of deaths or/and injuries. The key to lowering the death rate and damages to the health of collision accident victims is a timely arrival of Emergency Services to the accident scene. In the paper, we present and discuss the first results of the design and implementation of the vehicles collision detection system, which is based on a motion detector (MD) and Artificial Neural Network (ANN). To test MD and ANN separately, a small set of video records from traffic cameras that was not a part of a training dataset, were used. We found that while MD demonstrates reasonable performance, Haar Cascades-based pre-trained ANN requires significant improvements. Possible solutions to the aforementioned problem were proposed and discussed.

1 Introduction

Motor vehicle collisions often cause human deaths or health injury. The number of road traffic deaths continues to climb, reaching 1.35 million in 2016 [1]. The key point to shorten the death rate and health injury is the in-time arrival of Emergency Services to the traffic accidents. The effectiveness of the first aid rendering depends strongly on the ambulance arrival time which requires in-time Emergency Services alerting.

Local Russian regularities require all new cars to be equipped with ERA-GLONASS system that reports GPS (Global Positioning System) and/or GLONASS (Global Navigational Satellite System) coordinates in a case of an accident and provides emergency communications (“emergency button”) with Emergence Services. The system is not fully operational nowadays and relies on in-the-car sensors as well as the whole system proper functioning. The system is absent in a significant amount of older cars and may not function under certain conditions.

There is a “hype” nowadays on the so-called “smart-city” which is built around a number of so-called “smart” technologies. In any case, a number of video cameras, specifically, traffic cameras significantly increased during last decade. In practice, traffic cameras are generally used to detect traffic offenses but not to detect and alert on motor vehicle collisions.

In this paper we present the initial stage of our research on design and implementation of the system that should recognize and classify motor vehicle collisions basing on the traffic cameras video-streams analysis.

2 Methods

We are aimed at building a motor vehicles collisions detection system on top of a joint model consisting of an artificial neural network (ANN) and a motion detector (MD). Such models are often called as ensemble models. The key idea behind our model is to detect changes in dynamic characteristics of a video stream and to filter out those not related to motor vehicles. Considered in this paper approach relies on (1) the motion detector to detect changed areas in the frame and (2) the ANN to recognize motor vehicles in the same frame.

A widely used motion detection algorithm – known as the frame difference method – performs comparison of two or more sequential videoframes to detect the areas being changed. There are modifications that also perform background subtraction, different kinds of filtering, etc. (see, e.g., [2-4]). In our investigation we utilize a modification of the simple motion detection method. According to [5], motion detection is split into calibration and detection stages. The calibration stage requires in-pixel values comparison with the averaged RGB components. The detection stage requires estimation of the total measure of differences (in relative units, %) being detected. Our implementation labels the regions that exceed the averaged RGB components’ values in each frame, and then performs byte-wise comparison for two sequential frames. We also “hard-coded” the threshold value to define the significance of the differences.

To design an ensemble model, we use pre-trained Haar Cascades-based Artificial Neural Network (ANN)

* e-mail: y-romanovskaya@yandex.ru

[6] to recognize a vehicle on the given videoframe. According to the description [6], it was trained with 526 images of cars from the rear (360 x 240 pixels, no scale) as studying dataset. More precisely, the used ANN realizes Viola-Jones [7] algorithm which has the following pipeline: (1) Haar features selection; (2) construction of an integral image; (3) Adaboost training, and (4) cascading classifiers training.

At the early stages of the research it worth to investigate properties of the chosen algorithms manually. So, we selected a small set (10 videos) to test separately constituents of the constructed model – the motion detector and the ANN. A validation was performed visually by one of the co-authors. This set is not large enough to make any statistically significant validation of the model’s properties but it allows to roughly estimate and to illustrate regular features of the chosen algorithms. Table 1 enumerates the sources of the traffic cameras videos.

Table 1. Traffic cameras videos used to test ANN & MD

N	Description	URL
1	Dan Singha, Kathmandu Nepal Earthquake Traffic CCTV caught on camera April 25 2015	https://www.youtube.com/watch?v=btc020siGBo
2	Supercircuits, 3S N1011 5.0 Megapixel Indoor Network IP Box Camera: Highway Scene	https://www.youtube.com/watch?v=aio9g9_xVio
3	Tech Channel00001, Night Time Traffic Camera video	https://www.youtube.com/watch?v=G-ie5hQbG2s
4	SeattleTrafficCams !, Seattle Drivers vs the Snow - Round 2!	https://www.youtube.com/watch?v=Lgo6gG5Nn98
5	Asad Khan, Traffic camera footage from Bellevue	https://www.youtube.com/watch?v=ZwTUpQA6iwE
6	Sam Channel, Traffic Camera Footage of Fatal Crash	https://www.youtube.com/watch?v=IvarZkLc-Dc
7	Daniel-Catalin Stoica, Traffic camera in Russia	https://www.youtube.com/watch?v=cCtRWvZer4
8	Randy Pike, Singapore Road Traffic, CTE Tollroad Singapore	https://www.youtube.com/watch?v=IbYKYpqVEmw
9	bitgood, A3 A-road Traffic UK HD - rush hour - British Highway traffic May 2017	https://www.youtube.com/watch?v=UM0hX7nomi8
10	StormChasingVideo, 7/3/1999 Full Raw Traffic Camera Catching Explosive Storm Sewer Flooding - Full video	https://www.youtube.com/watch?v=Jp7zLbBs0Rc

3 Results

Results of both the motion detector and the Artificial Neural Network (ANN) application to the test set are illustrated in Figures 1-10. Figures 1-3 present specific features of the implemented simplified motion

detection algorithm. Figures 4-10 illustrate typical problems of the ANN.



Fig. 1. Illustration of the motion detector algorithm. Vehicles collision is detected (marked with rectangle).



Fig. 2. Illustration of the motion detector algorithm. Large ‘moving’ area is detected (marked with rectangle).



Fig. 3. Illustration of the motion detector algorithm. A few ‘moving’ areas are detected (marked with rectangle).



Fig. 4. Illustration of the ANN algorithm. Nothing is detected although there are vehicles to recognize.



Fig. 5. The same as Figure 4 but for another frame.



Fig. 6. Illustration of the ANN algorithm. Road defects are recognized (small squares) but not the vehicles.



Fig. 7. The same as Fig. 6 but for another frame containing vehicles collision.



Fig. 8. Illustration of the ANN algorithm. The shadow is recognized instead of vehicles.



Fig. 9. Illustration of the ANN algorithm. A side mirror is recognized instead of a whole vehicle.



Fig. 10. Illustration of the ANN algorithm. A taxi is detected due to special sign but not as a vehicle.

4 Discussion

We are aimed at building the motor vehicle collisions' detection system. The usage of AI technologies should significantly reduce the need for man-power to implement ad-hock algorithms and cost of them to maintain. It also should help in more rapid prototyping, data assimilation and eventually improve the accuracy, cost to develop and maintain, and scalability of the motor vehicle collisions' detection system. Good companions are computer vision (image recognition) algorithms (see, e.g., OpenCV library, etc.) in combination with pattern matching, statistical (correlation, regression) approaches as well as dataset pre-processing (normalization) and filtering to reduce the impact of noise. The latter is required in any cases, either with or without AI usage.

Our system is designed on top of an ensemble model which consists of an artificial neural network (ANN) and a motion detector. The model is expected to take traffic cameras' videostreams as input and classify it in the following categories: the probability of the traffic accident, severity of the injury basing on the class and properties of the accident. The preliminary classes might be car-to-human, multiple cars or car-to car (taking into account different types of cars, e.g., trucks, buses, cars, bikes, etc.). In this paper we concentrate on the initial stage, i.e. on the motor vehicles detection only.

Successful implementation of a detection system of such kind has to satisfy the following criteria: (1) true positives rate, i.e., traffic accident detection rate, should be of a value reasonable for practical purposes. True negatives are also important. (2) False positives rate, i.e., detections when no accident happened, should be below the threshold to make the detection practically reasonable. Too high false positives rate makes the solution unacceptable. False negatives are also in consideration but have no such a critical meaning. The solution also should (3) provide reasonable detection performance (ideally, near-realtime) on modern hardware and (4) be of comparable or a better quality over the existing alternatives.

The Results section demonstrates both false-negatives and false-positive cases. It is clear that the chosen solution is inappropriate for regular in-production use.

Figure 2 illustrates motion detector "misses". But it should be noted that motion detector itself operate as expected. It just detects corresponding pixels' values changes on two frames regardless of the nature of those changes. If desired, motion detector can be tuned by specifying the threshold and / or spatial area scale, or by implementing a full version of the algorithm. Some pre-processing of the video frames might be also desired to reduce the noise signal and improve the resulting detection.

The least well-done part of our model is an ANN. We expect this is due to the training dataset bias. The dataset should represent the real properties of the data to model. There are biases related to different characteristics of the traffic cameras / stream flows.

Some regularization (of resolution, frame-rate, etc.) is required to homogenize the properties of the stream. The view angle of a camera does matter as well, so some binning basing on the view angle or some data filtering is required.

Related to local conditions biases are also possible: most popular cars, road marking, traffic signs, driving conditions, weather conditions, light conditions, etc. Those differences should be reflected in filtering and / or binning procedures.

We see a few ways to improve the solution. We can try (1) to re-train the used ANN, (2) to search for and test another ANN, (3) to include another models into our ensemble model, or (4) to train the ANN ourselves.

An attempt to re-train the ANN does not look very promising. The ANN seems to be over-trained on specific taxi features; and over-trained ANN is hard to deal with. To search for another ANN that fits our specific task requirements is also an uncertain task with no guarantee that there are any. Extension of an ensemble model with a few specific models is in principle possible and is a widely used approach that requires proper specification of constitutes. For example, we could incorporate an object tracking model like in [8], an action recognition model by the analogy with [9], or to try dynamically define the detection threshold by the analogy with L-metrics from paper [10]. Training the model ourselves actually splits the task into two conjugated problems: (a) to define the architecture of the ANN, and (b) to prepare and validate the required datasets. The ANN's architecture should allow to save and transfer its state as well as to support a re-training ability. As for the dataset, there are two ways also: (a) to collect and markup data ourselves; (b) to reuse existing datasets.

Creation of new datasets requires the following questions to be addressed. ANN training requires significant labelled datasets of traffic cameras video streams (and partially photo if consider each single frame as separate image). Some metadata are also required including, but not limited to, co-ordinates (position), daytime (light conditions), season / weather or driving conditions (rain, snow, ice, etc.). Additional metadata and features are the subject for specific evaluation before making data mark up. Typical items to consider for marking up are traffic signs, cars, etc. Due to the nature of video streams data, the whole volume of the traffic camera streaming and traffic accidents video is orders larger over the amount we can store and perform robust mark up. Traffic camera streaming is provided in real time without the ability to obtain previous records if not saved previously. Publically available videos are updated eventually.

Thus, we have to define an optimal data size and to organize the storage and labelling (tagging) infrastructure. The datasets and model's lifecycle management is also required. Without properly organized maintain, the dataset and the model will soon become outdated (due to possible defects in labelling and the lack of new data as well as new labelling).

One possible solution is to incorporate it into educational process as part of the bachelor and master-degree courses. To build a public community around the dataset and the model is another option. But it will raise specific legal issues related to the community and development management, management for video-data and metadata rights, deposit of each active community member to the dataset, re-use the dataset for commercial, academic, and non-commercial purposes, etc... In any case, a communication with the interested parties as well as data dissemination are required.

5 Conclusion

As the result of the presented case study, we constructed the ensemble model (that consisted of a Motion Detector and an Artificial Neural Network) and evaluated it against the small dataset. Motion detector revealed reasonable performance but might be improved if required. The chosen implementation of the pre-trained ANN did not fit our constraints. It often detected small-scale details like side mirrors, taxi signs, wheels, etc... Thus, the ANN seems to be over-trained and requires significant improvements. For that reason, our future research implies (1) gathering and labelling of an appropriate dataset, and (2) training of the ANN ourselves.

References

1. World Health Organisation, 424 (2018), URL: <https://apps.who.int/iris/bitstream/handle/10665/276462/9789241565684-eng.pdf>
2. T. Yokoyama, T. Iwasaki, T. Watanabe, 7th Int. Workshop on Content-Based Multimedia Indexing, 201-206 (2009)
3. X. Zheng, Y. Zhao, N. Li, H. Wu, Int. Conf. Embed. Software and Syst., 541-543 (2009)
4. Y. Yong, R. Sudirman, K. Chew, 3^d Int. Conf. Comp. Intelligence, Modelling and Simulation, 46-50 (2011)
5. Widyawan, M.I. Zul, L.E. Nugroho, 9th Int. Conf. Ubiquitous Robots Ambient Intelligence, (2012) doi: 10.1109/URAI.2012.6462984
6. A.C. Sobral, (2016) https://github.com/andrewssobral/vehicle_detection_harcascades
7. P. Viola, M. Jones, Proc. IEEE Comp. Society Conf. Comp. Vision and Pattern Recognition, I-I (2001) doi: 10.1109/CVPR.2001.990517
8. W. Feng, Y. Cen, X. Zeng, Z. Li, M. Zeng, V. Voronin, KSII Trans. Internet Information Syst. **11**(11), 5459-5473 (2017)
9. V. Voronin, M. Pismenskova, A. Zelensky, Y. Cen, A. Nadykto, K. Egiazarian, Proc. SPIE 10802, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II, 1080200 (2018)
10. O. V. Zolotov, M. A. Knyazeva, Yu. V. Romanovskaya, Russian J. Phys. Chem. B **13**(4), 681-684 (2019)