Previous Research

（Multimodal Learning Models and Data Fusion based Techniques）

Multimodal Learning Models based on Data Fusion Analysis for Fully Autonomous Vehicle Navigation and Operation

June 2019

M.S. Student：SIMING ZHENG

Computer Graphics, Vision and Visualization (CGV2) Research Group
Augmented Reality Research Group
Department of Multimedia
Faulty of Computer Science, FCSIT
University of Putra Malaysia

Multimodal Learning Models based on Data Fusion Analysis for Fully Autonomous Vehicle Navigation and Operation

Keywords: On-board Sensors, Multi-target detection, Image Analysis and Processing, Instance Segmentation, Machine learning, Deep Learning Network, Cognition System, XGBoost Learning Framework, Data Fusion, Regression Trees, Decision-making.

1. Background and Motivation

Data fusion is one of the main applications of data preprocessing. It is also the primary research direction of Feature Engineering [1]. Different data fusion definitions exist in many scientific literatures. The Joint Directors of Laboratories (JDL) defines data fusion as a multi-level, multi-faceted process that automatically detects, correlates, estimates, and combines information and data from multiple sources [2].

In recent decades, many well-known researchers have applied data fusion technology to the research of autonomous vehicle [3] [4] [5]. It plays an important role in coordinating data fusion from multi-information sources, multi-platforms, and multi-user systems, and ensure the connectivity and timeliness of each sensor in the data processing system. The vehicle system can effectively understand the essential meaning of deep future spatiotemporal information, which describes the target location, movement, and its intentions. This technique provides sufficient and support information for autonomous vehicle in Level 3 and makes comprehensive automatic decisions in the Level 5. The data fusion was applied in many fields, such as signal processing, information theory, statistical estimation, reasoning, and artificial intelligence researches (machine learning and deep learning-based applications) [6][7]. As a consequence, we propose a Machine Learning and Deep Learning based learning framework, and combining the fusion data (Cameras, LIDARs, Radars, and other Fused Sensors) to forecast the spatiotemporal change information.

2. Research Objectives

At present, artificial intelligence algorithms are developing rapidly, and these new technologies play a more critical role in computer vision. Similarly, data fusion is widely used for intelligent video surveillance, man-machine interaction, automatic driving, and so on. The processing of data fusion is also a good basis for decision-making. In the next two phases (Leve 3 to Level 5), the learning model analyzes and predicts the fusion data, which is to make a judgment on the overall data and predict the subsequent operations to solve the practical problems of autonomous vehicle in our life [8]. All sensor data is transformed into structured data by the data fusion manner for the final decision-making in the control system. This proposal focuses on research objectives through the following six aspects:

1) Positioning and judgment operations for the pavement targets
Through the clustering analysis of the real-time monitoring data of the road surface, the target area of the road surface is found. Be quickly to check whether the area is human, animal, or moving object, analyze the recorded data and report to the relevant decision-making components.

2) Correlation analysis
Compare and analyze different types of data to find out if they are related. For example, how do real-time video and sensor data affect each other.

3) Dynamic change and road surface prediction
In order to predict the dynamic change of the autonomous vehicle, the most influencing factors should be analyzed as comprehensively as possible, and a real-time data of the road surface change should be recorded, such as a video or picture sequences.

4) Detection of incorrect data
By comparing real-time data with forecast data. If there is a big difference between them, or if the real-time data shows a jump, then it would indicate whether there is a problem with the monitoring sensor or whether there is an unusual situation to monitor. The cognition system can report problems and analyze the causes.

5) Advance warning of road conditions
This function is to estimate the behavior and direction of various road targets based on road conditions and warn of possible threats. For example, a fast-moving football on the highway, or a car that suddenly performs a U-turn, which are relative to the action detection and recognition researches.

6) The decision-making of autonomous driving
The decision classifier is used to make predictions, recommendations and driving operations based on road conditions, which finds the best decision for allowing the car to complete and continue the next operation. At the same time, the monitoring system is to make corrections or updates based on real-time road data.

3. Methodology and Algorithms

Base on the data fusion level, the processing methods can be divided into three levels: **1).** data level, **2).** feature level and the **3).** decision level. In this research, we also focus on the Feature Level and the Decision Level for improving the Automation Level from conditional automation (level 3) to full automation (level 5) [21] [22].

First, feature level includes the parameter-based classification and cognitive-based model according to the different measurements. Currently, the best solution to autonomous driving is parameter-based classification methods, which cover the research field of the statistical method and information technology [9]. Among those methods, machine learning-based Bayesian inference and deep learning-based convolutional neural network (CNN) are representative. Likewise, decision level-based data fusion also adopts machine learning and artificial neural network methodologies for the making-decision of autonomous driving [20].

Table 1. indicates different methodologies on the different component in cognition system.

Component	Methodology	Functions
For unstructured data: Customized Mask R-CNN Architecture	Region Proposal Networks （Region of Interest）	Detection / Classification
	Coordinate Transformation	Segmentation / Adaptive cruise control
	Temporal Correlation	Target Tracking
For structured data: XGBoost Computing Framework	XGBoost regression	Prediction
	Boosting	Decision-making for autonomous driving

Finally, this research attempts to propose a multimodal method based on deep learning and statistical machine learning to integrate a learning framework by using data fusion, which achieves a flexible and stable autonomous vehicle monitoring system. Contributions and improvements of this proposed research are expected as:

The learning model for the unstructured data (Images, Current frame, Video streaming from cameras):

A. Adopting a deep learning-based learning model is more practical and efficient to deploy in the automatic driving system, this also can refer to my previous research works based on the deep learning approaches [10] [11]. The Mask R-CNN is a multi-task learning architecture, which can implement different computer vision applications, such as classification, segmentation, detection on the current frame. Most importantly, the improved Mask R-CNN architecture can detect and identify multiple target objects for further segmentation and classification operations.

B. How to improve the Mask R-CNN model execute quickly on the autonomous driving system, which is also one of the innovations of this research [12]. This study will use the quantitative method to quantify the learning model to generate a stable and fast learning model. Recently, many neural network models have been put into practical use. The computational training demand grows linearly with the number of target objects, whereas the time period required for prediction is directly proportional to the number of targets, this means that the calculation efficiency becomes a crucial issue. The quantified learning model allows that hardware calculations simplify multiplication into simple accumulation operations as well as significantly reduce storage space in the mobile terminal. Quantifying the learning model can help device run the learning model faster and consume less power when executed on mobile computing devices, such as Google Tensor Processing Unit (TPU) and Nvidia Jetson Nano.

The learning model for structured data (The continuous and variables data from on-car sensors [22]):

A. In the **Signal Processing**, we need to convert the decision result into a digital signal in the final step. We propose to use the TXT or CSR data format in XGBoost[13] for the decision-making. Table 1 illustrates a summary of the most relevant options and training parameter for the learning model.

Table 2. illustrates the collected data mapping from different on-board sensors like camera, LiDAR, Radar and other Sonars.

Frame ID	Cameras	LiDAR	Radar and Sonar
Information source	Information of objectives. （The number, size, and classifications, etc.）	3D coordinate data of points, density	Distance, velocity
Data structure	a:b:c:d:e	x, y, z: d	Meters: Mile/h
Video Frame From ID 1 to ID 100,000	3:2:3:4:5	3562023.324, 532633.113, 198.734: 80	28: 15 mile/h

Cameras: Cameras are mostly used for object **recognition** and object **tracking** tasks such as lane detection, traffic light detection, and pedestrian detection. We can use cameras to detect, recognize, and track objects in front of, behind, and on both sides of the vehicle [19].

LiDAR: LiDAR is used for mapping, localization, and obstacle avoidance. Due to its high accuracy, LiDAR can be used to produce HD maps, to **localize** a moving vehicle against HD maps, to detect obstacle ahead.

Radar and Sonar: The signal simulation of radar and sonar is mainly used to simulate the signal of the target and its condition. For example, the data generated by radar and sonar show the **distance** as well as **velocity** from the nearest object in front of the vehicle's path.

B. **XGBoost (Extreme Gradient Boosting)** is an ensemble learning framework based on the Gradient Tree Boosting method. Its principle is to achieve accurate classification through the iterative calculation of weak classifiers [14]. XGBoost uses a tree set model, which is a set of classification and regression trees. Gradient enhancement is to construct a new regression tree to maximize the negative correlation with the gradient of the loss function, and further enhancing the flexibility of the enhancement algorithm [15]. In this study, all the collected information of the sensor devices (camera, lidar, and Radar) will be stored in the CSR file, which forms into a series of structured data set for the regression algorithm-based XGBoost framework. The reconstructed XGBoost framework reads every line dataset from the CSR file for the model learning (Because the regression algorithm can fit the training data set, see the example data in last line of Table 1) and prediction. **The predicted result can be transferred into control signal and applied for automatic control system's operations**, such as starting engine, the brakes, and pre-tensioning the seatbelts.

Advantages:

1. The improved Mask R-CNN architecture can effectively detect multi-target objects with higher real-time in the video stream. (1). It has the ability to identify and observe multiple objects of different types in each frame, such as human faces, animals, and every motion. This structure uses good CNN's characteristics and performance of extraction and classification to achieve target detection by Region Proposal [16]. The improved algorithm can be divided into four steps: candidate region generation, feature extraction, segmentation, and classification. For the unstructured data, we can regard it as a computer vision research problem. (2). The quantifying of neural networks has become a research hotspot in recent years. To generate a more efficient network and can be deployed on mobile terminals, lightweight network design is mainly to design very simple but high-performance networks, such as the MobileNet neural network.

2. It has the following advantages in dealing with the prediction problem: (1) XGBoost uses a parallel method to build a regression tree, and the CPU/GPU core of the computer during training has a higher computing speed [17]. (2) XGBoost is a general-purpose supervised machine learning method that achieves high-precision prediction in many practical applications [18]. Its high accuracy can be attributed to machine learning theory - Several weak classifiers construct one strong classifier, which can achieve better performance. It means that multiple weak learners can produce stronger learning than a single better model [17]. And the automatic summarization can be generated by using the proposed framework. For the structured data, we can regard it as a computer vision research problem.

4. Deliverables

Contributions and improvements of this proposed research are expected as:

A. A fast and flexible Mask R-CNN based detection system can satisfy the need for real-time automatic detection.

B. To quantify the learning model can get a greater performance to perform multi-targets detect computation, which can increase the speed of 5-10 percentage.

C. Using XGBoost can reconstruct the regression tree for predicting the continuous (variables) data. Regression trees can also cluster data and find the internal structure of the data and discover hidden associations.

D. XGBoost offers interfaces for multiple languages, including C++ and Java. This provides a convenient interface for cognition system calls in automatic driving. For the decision data generated by the model, we can convert them into digital signals for connecting with the input and output hardware interface circuits and controlling autonomous vehicle.

I will develop the above learning model and offer so-called assisted and autonomous driving technologies, which improve safety for the monitoring system in this research works.

5. Conclusion

Automatic driving is potential to alleviate pain and to have a broad societal impact by reducing the number of accidents, deaths, and injuries from traffic accidents. The core concept of automatic driving is to detect and analysis the movement trend based on the fusion information from multi objectives, which are collected from on-car sensors like cameras, radars and LIDARs. The study mainly introduces the idea of using the proposed methodologies to implement automatic driving. Research on the combination method of deep learning and machine learning to implement autonomous vehicles cognition system. The improved architecture of Mask R-CNN is to detect, learn and predict the intent of the target object's moving. Furthermore, the reason we use Gradient Tree Boosting is that many different classifiers trying to predict the same target variable do better than any single predictor. Gradient Tree Boosting is ideal for learning and predicting raw data collected by on-car sensors. We can use the proposed framework to implement a flexible and low-power cognition system in autonomous vehicle, which effectively retain the high steady-state precision and improve generalization performance.

7. References

[1] Zdravevski, E., Lameski, P., Trajkovik, V., Kulakov, A., Chorbev, I., Goleva, R., Garcia, N. (2017). Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering. IEEE Access, 5, 5262–5280. doi:10.1109/access.2017.2684913

[2] Ben Ayed, S., Trichili, H., & Alimi, A. M. (2015). Data fusion architectures: A survey and comparison. 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA). doi:10.1109/isda.2015.7489238

[3] Faouzi, N.-E. E., & Klein, L. A. (2016). Data Fusion for ITS: Techniques and Research Needs. Transportation Research Procedia, 15, 495–512. doi:10.1016/j.trpro.2016.06.042

[4] Bi, X., Tan, B., Xu, Z., and Huang, L., "A New Method of Target Detection Based on Autonomous Radar and Camera Data Fusion," SAE Technical Paper 2017-01-1977, 2017, doi:10.4271/2017-01-1977

[5] X. Meng, H. Wang and B. Liu, “A Robust Vehicle Localization Approach Based on GNSS/IMU/DMI/LiDAR Sensor Fusion for Autonomous Vehicles”, Sensors (Basel), September 2017, vol. 17, doi: 10.3390/s17092140

[6] Ni-Bin Chang, K. B. (2018). Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing, CRC Press, Boca Raton, Florida.

[7] Ghamisi, P., Hofle, B., & Zhu, X. X. (2017). Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(6), 3011–3024. doi:10.1109/jstars.2016.2634863

[8] Huang, W., Kunfeng Wang, Yisheng Lv, & FengHua Zhu. (2016). Autonomous vehicles testing methods review. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). doi:10.1109/itsc.2016.7795548

[9] Qian, Y., Zhou, W., & Wang, C. (2018). Research on Multi-Source Data Fusion in the Field of Atmospheric Environmental Monitoring. 2018 13th International Conference on Computer Science & Education (ICCSE). doi:10.1109/iccse.2018.8468770

[10] Siming Zheng, Rahmita Wirza O.K. Rahmat, Fatimah Khalid, Nurul Amelina Nasharuddin. 3D texture-based face recognition system using fine-tuned deep residual networks. 2019 Zheng et al. Published. doi: 10.7717/peerj-cs.236

[11] Siming Zheng, Rahmita Wirza O.K. Rahmat, Fatimah Khalid, Nurul Amelina Nasharuddin. A robust Iris Authentication System on GPU-Based Edge Devices using Multi-Modalities learning Model. (Under the peer review in Image and Vision Computing Journal. Cite as arXiv:1912.00756, [v1] Mon, 2 Dec 2019.)

[12] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN.The IEEE. International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969.

[13] Fangcheng Fu, Jiawei Jiang, Yingxia Shao, and Bin Cui. 2019. An Experimental Evaluation of Large Scale GBDT Systems. PVLDB (2019).

[14] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794. ACM, 2016.

[15] D. Nielsen, “Tree Boosting With XGBoost - Why Does XGBoost Win ‘Every’ Machine Learning Competition?,” 2016.

[16] Girshick, R., Donahue, J., Darrell, T., et al. (2016) Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 142-158. doi:10.1109/TPAMI.2015.2437384

[17] João, N. and Rui, F.N. (2019) Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to Trade in the Financial Markets. Expert Systems with Applications, 125, 181-194. doi:10.1016/j.eswa.2019.01.083

[18] Andreja, S., Nenad, S., Gordana, V., et al. (2019) Explainable Ex-treme Gradient Boosting Tree-Based Prediction of Toluene, Ethylbenzene and Xylene Wet Deposition. Science of the To-tal Environment, 653, 140-147. doi:10.1016/j.scitotenv.2018.10.368

[19] Weitang Song, Ge Zhang. Research and Analysis on Multi-sensor Data Fusion Algorithm Based on Intelligent Vehicle[J]. Modern Transportation Technology, 2012(3): 82-85.

[20] Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proceedings of the IEEE, 103(9), 1449–1477. doi:10.1109/jproc.2015.2460697

[21] Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., & Benediktsson, J. A. (2015). Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing. Proceedings of the IEEE, 103(9), 1585–1601. doi:10.1109/jproc.2015.2462751

[22] Sorber, L., Van Barel, M., & De Lathauwer, L. (2015). Structured Data Fusion. IEEE Journal of Selected Topics in Signal Processing, 9(4), 586–600. doi:10.1109/jstsp.2015.2400415