Traffic Sensory Data Classification by Quantifying Scenario Complexity*

Jiajie Wang[1], Chi Zhang[2], Yuehu Liu[3], Qilin Zhang[4]

Abstract— **For unmanned ground vehicle (UGV) off-line testing and performance evaluation, massive amount of traffic scenario data is often required. The annotations in current off-line traffic sensory dataset typically include I) types of roadways II) scene types III) specific characteristics that are generally considered challenging for cognitive algorithms. While such annotations are helpful in manual selection of data, they are insufficient for comprehensive and quantitate measurement of per-roadway-segment scenario complexity. To resolve such limitations, we propose a traffic sensory data classification paradigm based on quantifying the scenario complexity for each roadway segment, where such quantification is jointly based on road semantic complexity and traffic element complexity. The road semantic complexity is a proposed measurement of the complexity incurred by the static elements such as curvy roads, intersections, merges and splits, which is predicted with a Support Vector Regression (SVR). The traffic element complexity is a measurement of complexity due to dynamic traffic elements, such as nearby vehicles and pedestrians. Experimental results and a case study verify the efficacy of the proposed method.**

I. Introduction

During the developing, testing and verification cycle of unmanned grounded vehicle (UGV) system (e.g. [12]), a large amount of traffic scenario data is utilized for performance evaluation. In recent years, to meet the practical demand of autonomous driving technology, especially for the research and development on environmental cognition and understanding algorithms [20]–[22], many traffic scene datasets have been proposed, such as KITTI [5], RobotCar [10]. These datasets are usually collected in traffic scenarios with dynamic changes in cognition complexity, including different types of roads, scene contents and scene characteristics. However, the lack of quantitative characterization of the scene complexity in these datasets could impede interpretable evaluation of UGV systems. On the other hand, in the unmanned off-line testing [2], [7], we find that there is usually a negative correlation between the unmanned vehicle algorithm performance and scenario complexity. Traffic data with higher scenario complexity typically leads to worse performance of the environment cognition and understanding algorithm. If we use unorganized data to test and evaluate an environment-aware understanding algorithm for an unmanned system, the results are mostly indistinguishable: an algorithm with 0.84 overall accuracy may consistently perform worse than a competing algorithm with only 0.83 overall accuracy in common road scene scenarios. Therefore, the complexity of the scene data needs to be incorporated for reliable evaluation of UGV systems.

Hence we propose a method in quantifying scenario complexity to rank massive scene data. The complexity is calculated on the basis of the road types, scene types, challenging condition and traffic elements. Scenario complexity is computed from two perceptual data levels: 1) *Road semantic complexity* (RSC). We propose a road semantic complexity prediction method based on support vector Regression (SVR). The road semantic complexity of a given non-hierarchical semantic descriptor is predicted by learning the relationship between the road label and the semantic descriptor. 2) *Traffic element complexity* (TEC). Traffic elements are moving entities that participate in road traffic activities. In this paper, vehicles are chosen to be the representative of the traffic elements. We devise description matrices of traffic elements and TEC calculation to quantify the complexity.

The contributions of this paper are as follows.

1) Traffic sensory data is semantically quantified in terms of scenario complexity.
2) A comprehensive scenario complexity is formulated based on scene types, test sites/location information and dynamic traffic elements on a per-segment basis.

This rest of the paper is organized as follows. Section II reviews existing datasets and their respective problems, and the formulation of scenario complexity. In the Section III, a new scene semantic feature is proposed, followed by the sensory data classification framework. Section IV introduces related applications, including accelerating off-line UGV evaluation and grading data synthesis. Section V summarizes and concludes the paper.

II. Related Work And System Overview

Several road-sensing datasets for unmanned vehicle testing have been proposed since 2012, including KITTI [4], [5], RobotCar [10], Cityscape, Udacity, BDDV, etc. The KITTI dataset was proposed by researchers from Karlsruhe Institute of Technology and the Toyota Institute of Technology at Chicago, which is the largest multi-sensory[1] traffic scene

*This work was supported by National Natural Science Foundation of China (Grant No: 91520301).

[1]Jiajie Wang is with the school of Software Engineering, Xi'an Jiaotong University, Xi'an, China.

[2]Chi Zhang is the corresponding author and is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China. colorzc@stu.xjtu.edu.cn

[3]Yuehu Liu is with the Faculty of the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China.

[4]Qilin Zhang is with HERE Technologies, Chicago, Illinois, 60606, USA.

[1]Including stereo RGB/grayscale cameras, LIDAR and GPS/IMU. Multi-sensory data, especially remote sensing LIDAR data [1], [18], [19], provides more discriminative information for object detection, classification and localization.

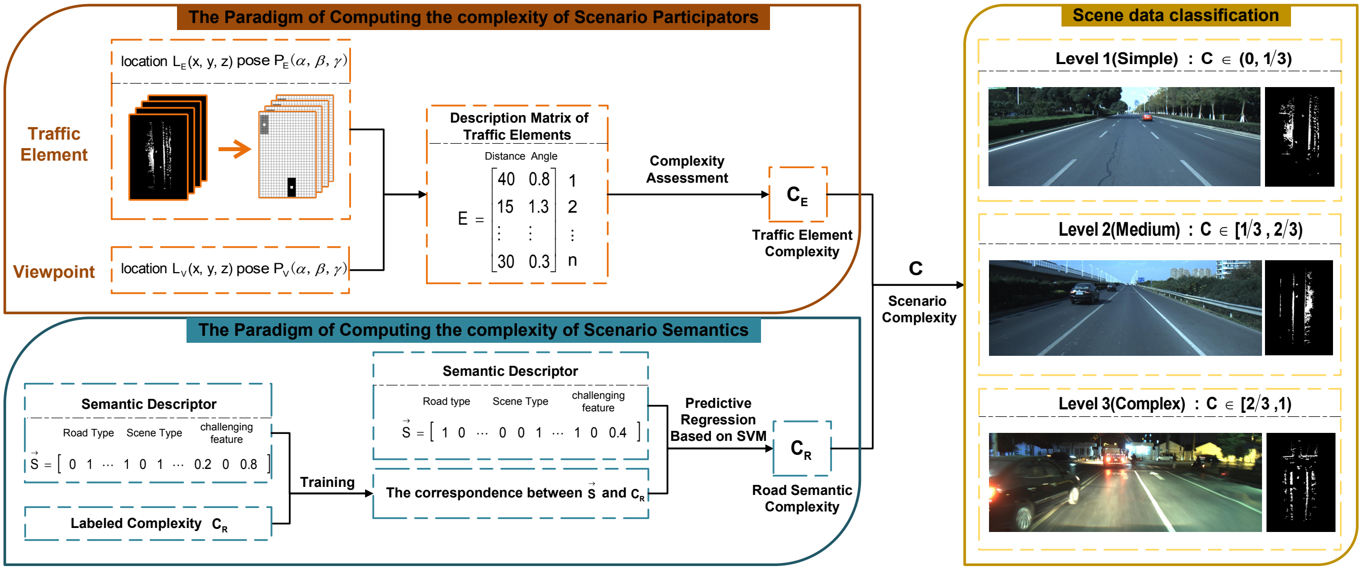


Fig. 1. The detailed illustration of the proposed traffic sensory data classification via quantifying scenario complexity.

dataset for autonomous driving. This dataset is used to evaluate the performance of various technologies such as visual odometry, object detection, and 3D tracking for UGVs. The RobotCar dataset presented by Oxford University contains 100 repetitive driving data for a fixed driving circuit in Oxford within a year. The dataset captures different weather, traffic and includes long-term changes in building and road construction.

These datasets describe the complexity of the road scenes collected in different degrees and at different levels. The KITTI raw datasets are divided into 'Road', 'City', 'Residential', 'Campus' and 'Person' by scene type. The label of 3D objects detecting is subdivided into car, van, truck, pedestrian, cyclist, tram, and misc. The dataset is initially classified based on the scene type, and the scene complexity is represented by describing the type and number of vehicles for each type of scene. However, simply categorizing the scene does not provide a detailed description of the scene features. Unlike KITTI, RobotCar describes complexity of the scenes under different conditions, including pedestrian, bicycle and vehicle traffic, light rain, heavy rain, direct sunlight, etc. Compared with KITTI, the classification of RobotCar presents more details, but it still lacks structured semantic descriptions. For example, they don't have a structured semantic description of the road types, and ignore the impact of traffic conditions on scenario complexity, such as overtaking and pedestrian avoiding, and so on.

In view of the shortcomings of the existing database, this paper describes a method of scene complexity calculation from the aspects of road types, scene types, challenging conditions and traffic elements. The complexity of the scene C is measured by the quantified road semantic complexity C_R and the traffic element complexity C_E. The description is as follows:

$$C = \lambda_1 C_R + \lambda_2 C_E \tag{1}$$

C_R is predicted by SVR based on the scene semantic features of artificial annotations. C_E is calculated according to the distance and angle between the dynamic obstacles and the viewpoint on the road.

Grading the traffic scene data is a challenging issue. The traditional method doesn't pay enough attention to grading scene data, while the current method is based on manual grading. However, massive scene data in autonomous vehicle testing relying on the manual classification is far from the time requirement. As shown in Figure 1, after calculating the complexity of the scene, we classify the scene data into three levels: general, medium, and extreme. This method is applied to the off-line testing of unmanned vehicles in November 2017. The fourth chapter of the article will introduce this application in detail.

III. Semantic Feature

Numerous traffic scene images provide data support for unmanned off-line testing. However, the existing database lacks the quantitative description of scene data complexity and scene characteristics. Therefore, we propose a method of data classification based on the quantization of the complexity of the scene. This quantitative complexity classification method is considered from two levels: I) which describes the semantic characteristics of roads through vectors. II) describes the topology information of traffic elements through matrix.

A. Quantizing The Description Of Road Scene

According to the intrinsic semantic attributes of Unmanned vehicles test site and the factors influencing autonomous vehicles safety and algorithm performance, we define the semantic descriptor from three levels: road types, scene types and challenging conditions. The scene described by the perceived data is described as follows:

Road types (RT) include urban areas, high speed, rural areas, and so on. The road types reflect the basic pattern of the scene, and different road types reflect different scene contents. In different scenarios, the categories and quantities of challenging characteristics are also different. For example, urban areas have many road junctions and pedestrians, and there will be a viaduct on the highway or a toll station. The road types are described by n-dimensional vector with a value of 0 or 1 which is used to determine the unique road type.

Scene types (ST) include normal driving, intersection, up/down viaduct, through charge, tunnel, turntable, steep slope, bridge, railway, etc. Scene types reflect the semantic content of the scene. Our proposed scenarios cover all scenes which the autonomous vehicles daily driving through, and the different scene types are independent of each other. The scene type is described by a m-dimensional vector with a value of 0 or 1 which is used to determine the unique scene type.

Challenging conditions (CC) is a challenging road factor for the environment cognition algorithm in the scene data frame which include bend, overtaking, pedestrian avoidance, construction, large car flow, haze, night, road surface traces, lane line blurred, light influence and so on. The number and extent of the test data in the scene data frame directly decide the complexity of the scene. The type of challenging conditions is described by the o-dimensional vector whose value is 0, 0.2, 0.4, 0.6, 0.8, 1 which represents the degree of the different challenging conditions.

On the basis of the semantic descriptor of the road scene at the above three levels, we get a M-dimensional ($M = m + n + o$) vector. According to the performance of all teams in *Intelligent Vehicle Future Challenge 2017* (IVFC 2017), We propose a method which is described as Eq. (2) to calculate $C_R^{(t)}$.

$$C_R^{(t)} = 1 - \sum_i^N \sum_j^{M'} \frac{F_{ij}^{(t)}}{NMmax(F_i^{(t)})} \tag{2}$$

M' represents the number of participating teams. N is the number of tasks in off-line testing of t-th IVFC. F (F1-Measure) is the harmonic average of accuracy and recall. The tasks in IVFC 2017 are lane keeping capacity assessment, the detection and identification of front vehicle and pedestrian [3], and basic traffic signal detection. $C_R^{(t)}$ within the interval $[0, 1]$ is obtained by normalization.

However, it is unrealistic to manually annotate all scene data in a dataset, due to the extremely time-consuming nature of such annotation. To automate the scene complexity assignment, we propose a machine learning-based approach. Inspired by recent advancement in supervised learning [6], [11], [16], SVR [4], [9] is exploited to account for the relationship between the road marking complexity and the semantic descriptor. We manually annotate a small subset of representative traffic scenario data for training the SVR, and the trained SVR is used to predict the complexity of the entire dataset. The average training time is 1.18 second per 100 samples on an Intel Core i5 7200U laptop while the average precision of predicted complexity on the training set and testing set is 93.23% and 68.72% respectively.

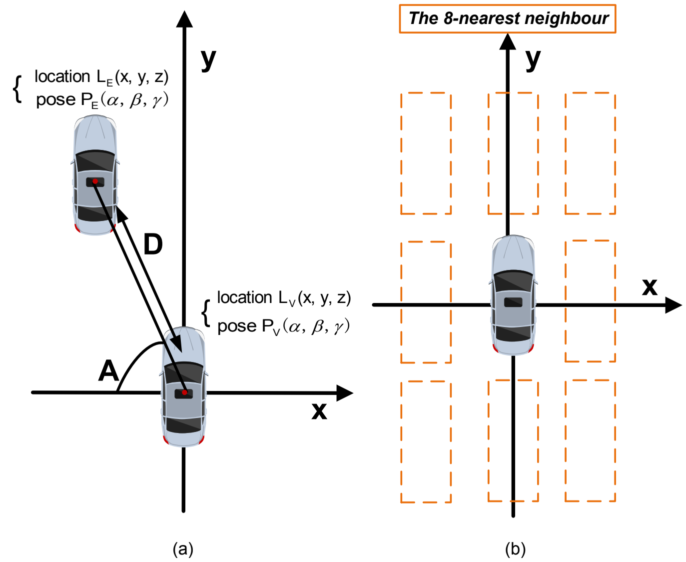


Fig. 2. A schematic diagram for calculating the complexity of traffic elements. (a) Calculating distance and angle from location and pose of traffic elements and viewpoint. (b) The 8-nearest neighbour of viewpoint.

B. *Quantizing The Description of Traffic Elements*

The challenge of driving unmanned vehicles on real roads is the interference of other traffic elements [15]. Taking unmanned vehicle as the viewpoint of driving, the distance and angle of other vehicles from unmanned vehicles have an impact on the difficulty of the performance testing of unmanned vehicles. Thus we define a description matrix of traffic element E which describes the traffic elements reflected by semantic data in two aspects: distance and angle.

Assuming that there are N vehicles on the road except for the unmanned vehicle, each row in the description matrix of traffic element represents each car's information: the first column means the distance from the car to the car; the second column stand for the car's viewpoint. The distance and angle of each traffic element from the viewpoint are calculated by the point cloud which is collected by LIDAR, then we obtain the traffic element matrix. According to the matrix, we have designed an assessment to calculate the complexity of traffic elements.

As shown in Fig.2, the N participators in the scenario can be described by a $N \times 2$ matrix, where the j-th row is denoted by a 2-parameter vector $\mathbf{s}_j = [D_j \;\; A_j]$ $(j = 1, 2, ..., N)$, where D_j denotes the distance between the geometric center O_j of j-th participator and the origin O of the autonomous vehicles. A_j is the minimum angle from O_jO to the x-axis of the body coordinate system. Based on this description matrix, the complexity C_{Ej} of j-th participator's contribution is the summation of the horizontal contribution $x_j = D_j \cos(A_j)$

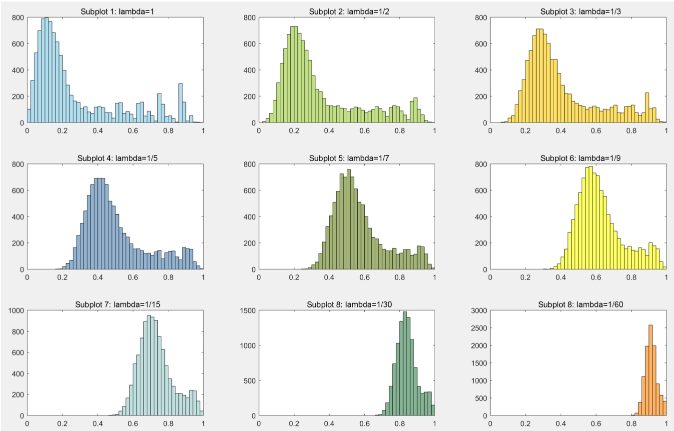


Fig. 3. Scene complexity distribution under different lambda.

and vertical contribution $y_j = D_j \sin(A_j)$:

$$C_{Ej} = \alpha e^{-(\lambda D_j \cos(A_j))} + \beta e^{-(\lambda D_j \sin(A_j))} \quad (3)$$

Where α and β are the weights of each direction of contribution, respectively. At present, we consider that the two contributions are the same as 0.5. Considering that the result of complexity is small, we pose a scaling factor λ, as shown in Fig.3. Assuming that such distribution is subjected to Gaussian distribution. When λ is equal to $1/7$, the mean of Gaussian distribution is 0.5, which is more reasonable. For a scenario of N traffic participators, not all the participators in the scenario make contribution to the overall complexity. For instance, in the worst situation, the acquisition platform is surrounded by 8 vehicles, which means the rest of the participators contributes rarely to the complexity, shown in Fig.3.(b). Such case illuminates us that the overall complexity could be calculated by the summation of the complexities of the 8-nearest participators, as in following equation:

$$C_E = \frac{1}{8} \sum_{i=1}^{N'} \alpha e^{-(\lambda D_n \cos(A_n))} + \beta e^{-(\lambda D_n \sin(A_n))} \quad (4)$$

where

$$\begin{cases} N' = N & N < 8 \\ N' = 8 & N \geqslant 8 \end{cases} \quad (5)$$

As shown in Eq. (5), N represents the total number of vehicles, N' represents the actual number of vehicles required for calculating complexity. When the total number of vehicles is more than 8 vehicles, it is only calculated according to the nearest eight vehicles. When the total number of vehicles is less than 8 vehicles, the actual number is calculated.

C. Verification

The purpose of the experiment is to verify the reliability of our proposed method and to demonstrate the necessity of quantifying C in unmanned vehicle off-line testing. The Fig.4 show the results of three experiments in three different complexities. We can find that the calculated complexity is consistent with the complexity of the scene. In Fig.5 and Fig.6, with the same C_R, the higher the complexity of traffic elements get the higher the traffic scene complexity under the same C_R. In the case of the same C_E, the higher the C_R obtains the higher the complexity.

In contrast to the bottom image of Fig.4-Fig.6, the semantic descriptor is almost the same, but why is the complexity of Fig.6 the highest? Because there are other vehicles around the unmanned vehicle from LIDAR data. It is difficult to see other vehicles parallel to the unmanned vehicle from the image data which requires LIDAR data to participate in the computation of complexity.

IV. Application

A. Hierarchical Accelerated Testing

The method of this paper classifies the mass traffic scene data based on complexity, and calculates the distribution of different scenes so as to accelerate the process of simulation test [23], [24] of the autonomous vehicle.

Take our off-line test task in November 2017 as an example. The algorithm performance of the unmanned vehicle is evaluated by the algorithm completed under the 2000 km road data. The test data are derived from the scene video data and the 3D laser point cloud [14] collected by the real road. It's a time-consuming task to complete 2000 kilometers of data. In order to speed up the off-line test process, we use the method of this paper to classify a large number of scene data. Road data are divided into three levels according to complexity: general data whose complexity is between 0 and $1/3$, medium data whose complexity is between $1/3$ and $2/3$ and extreme data whose complexity is between $2/3$ and 1. In hierarchical data, the approximate distribution of data at all levels is obtained according to statistical methods: simple data account for 89.29%, medium data account for 8.93%, and extreme data account for 1.78%.

According to the data distribution of three levels, the probability of occurrence of each level scenario is calculated to get the equivalent ratio. Thus, it helps reduce the road length of actual testing and improve the efficiency of off-line testing. As shown in Table 1, each 1 km of simple road data is equivalent to 1 km in real traffic environment; 1 km of moderately difficult road data is equivalent to 10 km in real traffic environment; and every extreme road data one kilometer is equivalent to driving 50 kilometers in real traffic environment.

TABLE I
Road Mileage And Equivalent Length Under Different Complexity

	simple	medium	complex
Urban Road	35Km	14Km	10.5Km
Suburbs Road	15Km	6Km	4.5Km
Highway	50Km	20Km	15Km
Total	100Km	40Km	30Km
Equivalent Length	100Km	400Km	1500Km

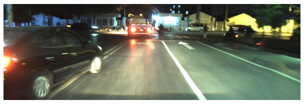
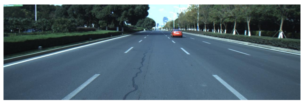
Traffic Scene	Point Cloud Data	Scene description vector	Traffic element description matrix	Scene complexity
		$\vec{s}=[1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.8\ 0.4\ 0\ 0.2\ 0\ 0.6\ 0.2\ 0\ 0.4]$	$\begin{bmatrix} 26.2837 & 0.0711 \\ 37.6226 & 0.2507 \\ 46.2316 & 0.3523 \\ 7.0120 & 0.3624 \\ 16.7612 & 0.3802 \end{bmatrix}$	**0.7924**
		$\vec{s}=[0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0\ 0.2\ 0\ 0\ 0.2]$	$\begin{bmatrix} 13.7306 & 0.2243 \\ 40.0248 & 0.1023 \end{bmatrix}$	**0.4531**
		$\vec{s}=[1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0]$	$\begin{bmatrix} 41.7721 & 0.0728 \end{bmatrix}$	**0.2086**

Fig. 4. Three kinds of scenes with different complexity.

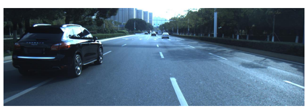
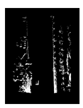
Traffic Scene	Point Cloud Data	Scene description vector	Traffic element description matrix	Scene complexity
		$\vec{s}=[1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0\ 0\ 0\ 0\ 0\ 0.4\ 0.2\ 0.2\ 0.8]$	$\begin{bmatrix} 7.7761 & 2.6703 \end{bmatrix}$	**0.5903**
		$\vec{s}=[1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0.2]$	$\begin{bmatrix} 10.5447 & 3.3420 \end{bmatrix}$	**0.2556**

Fig. 5. Comparison of scene complexity of different road semantic complexity under the same traffic elements.

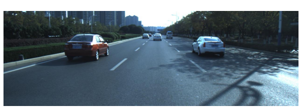
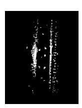
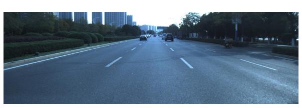
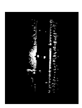
Traffic Scene	Point Cloud Data	Scene description vector	Traffic element description matrix	Scene complexity
		$\vec{s}=[1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0\ 0\ 0.2\ 0\ 0.2]$	$\begin{bmatrix} 14.6497 & 0.2481 \\ 23.2881 & 0.1782 \\ 24.9377 & 0.1244 \\ 4.0087 & 1.0449 \\ 18.5298 & 0.2284 \\ 47.6569 & 0.0212 \end{bmatrix}$	**0.5186**
		$\vec{s}=[1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.2\ 0\ 0\ 0\ 0.2\ 0\ 0.2]$	$\begin{bmatrix} 4.1286 & 1.0796 \\ 4.4082 & 1.4100 \end{bmatrix}$	**0.2956**

Fig. 6. Comparison of scene complexity of different traffic elements under the same road semantic complexity.

B. Grading Data Synthesis

After the dataset is graded, we can generate the scene according to the specified complexity. For the given complexity, the semantic descriptor and the traffic element description matrix which satisfy the complexity calculation formula are generated by random sampling. In traffic scene generation tool (dTraffic [17], Matlab automatic driving tool box, AirSim [13]), we construct traffic scenes that conform to semantic descriptor and traffic element description matrix, and generate scene images under given viewpoint.

Grading data synthesis can benefit from augmented reality (AR) technology by exploiting traffic scene videos as baselines, with additional traffic elements augmented on top of such videos. Examples include placing additional virtual vehicles on the pavements of base traffic scene videos [25]. After mapping the environment perception data and virtual reality registration, we simulate the corresponding pose between the camera and the real camera. Then the simulated sensor data obtained can correspond to the sensing data obtained by the real camera. At this time, a simulated camera at each moment is set up to obtain the simulated sensing data and fuse with the real video image data. The major steps are as follows. Firstly, the simulated camera is set up to observe the simulation road scene according to the camera internal parameters and position [8]. The simulation cameras in the road scene are set in sequence to the position and posture of the camera and the road scene is observed in the order of real image acquisition. Secondly, the virtual vehicles that have known coordinates are projected to the image plane. The transparency of the virtual road space and the sky environment is adjusted to zero, and the traffic elements are projected only to the image plane. Finally, the traffic element projection is fused with the real scene video image which is perceived by the current spatio-temporal position.

V. CONCLUSIONS

This paper proposes a new approach to quantize the complexity of traffic scene. In our method, the road semantic complexity is forecast based on SVR while the complexity of traffic elements is obtained by 8-nearest neighbor analysis. Further studies explain how the proposed quantitatively method measures scenario complexity accurately. This technique thus has the potential to grade for mass traffic data scientifically and speed up the simulation test of unmanned vehicles, verified by extensive applications.

REFERENCES

[1] H. Abeida, Q. Zhang, J. Li, and N. Merabtine. Iterative sparse asymptotic minimum variance based approaches for array processing. *IEEE Trans. Signal Process.*, 61(4):933–944, 2013.

[2] A. Diosi, S. Segvic, A. Remazeilles, and F. Chaumette. Experimental evaluation of autonomous driving based on visual memory and image-based visual servoing. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):870–883, 2011.

[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, April 2012.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[5] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, June 2012.

[6] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li. Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA*, 2-7 February, 2018.

[7] L. Li, W. L. Huang, Y. Liu, N. N. Zheng, and F. Y. Wang. Intelligence testing for autonomous vehicles: A new approach. *IEEE Transactions on Intelligent Vehicles*, 1(2):158–166, June 2016.

[8] Y. Li, Y. Liu, C. Zhang, D. Zhao, and N. Zheng. The "floor-wall" traffic scenes construction for unmanned vehicle simulation evaluation. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1726–1731, Oct 2014.

[9] C. Lopez-Martin, S. Banitaan, A. Garcia-Floriano, and C. Yanez-Marquez. Support vector regression for predicting the enhancement duration of software projects. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 562–567, Dec 2017.

[10] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 year, 1000 km: The oxford robotcar dataset. 36, 11 2016.

[11] L. Ran, Y. Zhang, W. Wei, and Q. Zhang. A hyperspectral image classification framework with spatial pixel pair features. *Sensors*, 17(10):2421, 2017.

[12] L. Ran, Y. Zhang, Q. Zhang, and T. Yang. Convolutional neural network-based robot navigation using uncalibrated spherical images. *Sensors*, 17(6):1341, 2017.

[13] S. Shah, A. Kapoor, D. Dey, and C. Lovett. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. *Field and Service Robotics*, pages 621–635, November 2017.

[14] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, pages 75–82, March 2014.

[15] J. Xie, M. Kiefel, M. T. Sun, and A. Geiger. Semantic instance annotation of street scenes by 3d to 2d label transfer. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3688–3697, June 2016.

[16] J. Zang, L. Wang, Z. Liu, Q. Zhang, Z. Niu, G. Hua, and N. Zheng. Attention-based temporal weighted convolutional neural network for action recognition. In *Proceedings of the 14th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations, AIAI 2018, Rhodes, Greece*, 2018.

[17] C. Zhang, Y. Liu, D. Zhao, and Y. Su. Roadview: A traffic scene simulator for autonomous vehicle simulation testing. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1160–1165, Oct 2014.

[18] Q. Zhang, H. Abeida, M. Xue, W. Rowe, and J. Li. Fast implementation of sparse iterative covariance-based estimation for source localization. *J. Acoust. Soc*, 131(2):1249–1259, 2012.

[19] Q. Zhang, H. Abeida, M. Xue, W. Rowe, and J. Li. Fast implementation of sparse iterative covariance-based estimation for array processing. In *Proceedings of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA*, pages 2031–2035, 6-9 November, 2011.

[20] Q. Zhang and G. Hua. Multi-view visual recognition of imperfect testing data. In *Proceedings of the ACM International Conference on Multimedia, Brisbane, Australia*, pages 561–570, 26-30 October, 2015.

[21] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang. Can visual recognition benefit from auxiliary information in training? In *Proceedings of the Asian Conference on Computer Vision, Singapore*, pages 65–80, 1-5 November, 2014.

[22] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang. Auxiliary training information assisted visual recognition. *IPSJ Trans. Comput. Vis. and Appl.*, 7:138–150, 2015.

[23] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. LeBlanc. Accelerated evaluation of automated vehicles in car-following maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, PP(99):1–12, 2017.

[24] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Transactions on Intelligent Transportation Systems*, 18(3):595–607, March 2017.

[25] D. Zhao, Y. Liu, C. Zhang, and Y. Li. Autonomous driving simulation for unmanned vehicles. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 185–190, Jan 2015.