3D Traffic Scenes Construction and Simulation based on Scene Stages

1st Jian Yuan
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China

2nd Yaochen Li*
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China

3rd Huayun Pan
School of Software Engineering
Xi'an Jiaotong University
Xi'an, China

4th Zhichao Cui
Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University
Xi'an, China

5th Yuehu Liu
Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University
Xi'an, China

***Abstract*—In this paper, we present a method for 3D traffic Scenes Construction and Simulation based on scene stages. The construction method is divided into two phases: scene stages analysis phase and scene model construction phase. During the analysis phase of the scene stages, the content of the road image is analyzed by using semantic labels, then we define road scene elements based on the graph models. Furthermore, the wireframe models of traffic scenes are constructed. During the phrase of scene model construction, the scene layout is firstly displayed based on the Scene Stages, and then the road scene model of the 3D corridor structure is quickly constructed. The applications for road scene simulation are then developed, which prove the effectiveness of the proposed framework.**

***Index Terms*—Scene stage, scene construction, semantic label, graph model**

I. Introduction

With the development of artificial intelligence technology, unmanned vehicle has been a hot research topic. At present, the advanced driver assistance system (ADAS) has been widely studied in the community of intelligent transportation systems [11]. Autonomous driving will greatly reduce the accident rate caused by human operations, and it will also improve the traffic congestion. In the field of unmanned vehicles, the development of technology must be based on the premise of human safety. In order to test the safety level of autonomous driving, it may require tens of thousands of miles or even billions of miles of experimental milestones to prove it. The unmanned system uses lasers, cameras, and other devices to collect road datasets. The driver-free perception algorithms of computer vision can be promoted, and the datasets can be used to verify the algorithm.

This work was supported by the National Natural Science Foundation of China under Grant 61803298, Natural Science Foundation of Jiangsu Province under Grant BK20180236. Yaochen Li is the corresponding author: yaochenli@mail.xjtu.edu.cn

Virtual simulation testing is an important part for the test of unmanned vehicles. Virtual simulation provides test scenarios for unmanned vehicles, such as object detection and tracking, and visual mileage calculations. 3D traffic scenes construction based on 2D images is an important part of virtual simulation. Based on the collected datasets, the scene is reproduced in the virtual environment without the limitations of the real environment. 3D traffic scene construction can effectively simulate the real traffic environment, and provide a test environment for of unmanned vehicles. For example, Google developed the virtual simulation system which is named "Carcraft" for driverless testing. The testers use Carcraft to construct a large number of scenarios, allowing the autopilot program to undergo rounds of practice and improve its capabilities.

We propose a method for 3D traffic scenes construction and simulation based on scene stages. In the following sections, the modeling method and experiments will be described in detail.

II. Related Works

The data acquired directly from two-dimensional images will lose certain information relative to real three-dimensional scenes. Therefore, how to reconstruct three-dimensional scenes using 2D images becomes a hot issue. The method to extract 3D layouts from 2D images has been extensively studied [1]. Humans can easily grasp the spatial layout of scenes displayed by the 2D images, and they can make estimate of the entire space [2]. Many methods have been proposed to recover the geometry of a single image. Hoiem et al. [5] [6] provide a multiple-hypothesis framework for reliably estimating the scene structure from a single image. According to Hoiem's method, each image area is divided into three categories: "support", "vertical" and "sky". By modeling the powerful clues of color and texture, we can implicitly identify the objects that correspond to a particular geometry class. Nedović

et al. [7] think that scene stages have many advantages. Stages can provide viewers with information about the semantic backgrounds, the identity of scene elements, etc. They propose a method for inferring weak scene set from a single image and building a rough sketch of the scene model. Lou et al. [8] propose a framework for inferring pixel-level 3D layouts by using a global image structure, they use image segmentation to obtain stage classifications and generat 3D layouts by stage classification.

For restoring real 3D scenes, Google Street View [9] uses a 360-degree panoramic view of a location to project a panoramic photo onto a spherical 2D surface. Google Street View supports a unique 3D navigation. In this mode, the user clicks the mouse at a certain point in the scene, and the 3D navigation conveys the user's perspective to the image closest to the point's 3D position [10]. The method of Microsoft Street Slide [12] is another example. Microsoft Street Slide connects the photos taken as a long strip, providing a single location panoramic view and street-level view mode, providing convenience for users to browse the streets.

The rest of this paper is organized as follows: In Section III, we propose two types of scene models according to the complexities of input road image sequences. The simulation process of the simple traffic scenes is described in Section IV. For more complex image sequences with foreground objects, the cluttered traffic scenes are simulated in Section V. Experiments and comparisons are shown in Section VI, followed by the user study performed in Section VII. Finally, we close this paper with the conclusion and future works.

III. Semantic Labeling Method and Graph Model

A. Usage of Semantic Labeling Method in Traffic scenes

Visual scene understanding is one of the challenges in the computer vision community, which aims to provide detailed information of objects in the scene. In this paper, we use semantic annotation method to infer the traffic scene information. With the help of InteractLabeler tool, we performed pyramid segmentation of traffic scene images, and then perform semantic labeling based on the segmentation results.

Pyramid segmentation, as the name implies, aims to segment images like a pyramid. Layers of images make up the pyramid, and the images from the bottom to the top gradually decrease in size and resolution. Therefore, the resolution of the image to be processed at the bottom of pyramid is the highest, and that of image to be processed at the top of pyramid is the lowest. The maximum layer of the pyramid is half of the larger one in the height and width of the image, so the height and width of the image to be processed must be integer multiples of 2. In addition, the lower layer is sampled to obtain the upper layer, so the upper and lower layers of the pyramid form a mapping relationship.

B. Graph Models of Traffic Scenes

We can easily infer the semantic labels. In the traffic scene, there are a large number of background elements which make up the traffic scene, such as the sky, road surface, and both

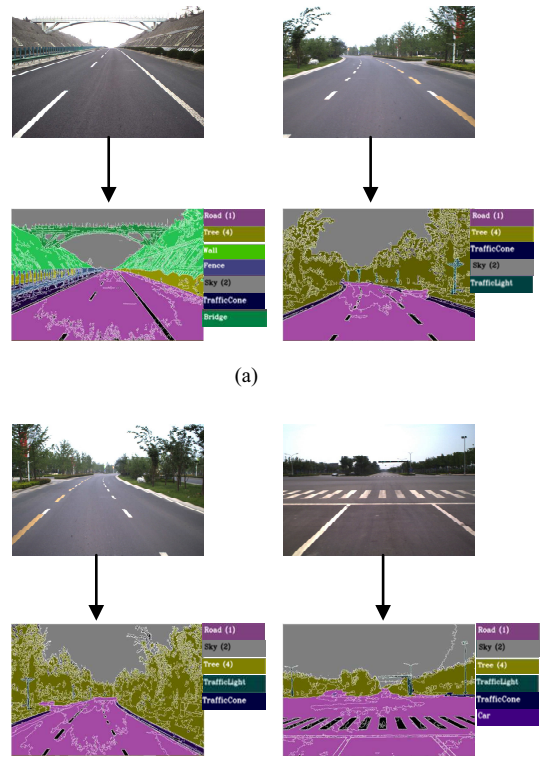

Fig. 1. Semantic labeling of the traffic scenes.

sides of the wall. In addition to the background elements, foreground elements are also part of the traffic scene, which includes traffic lights, traffic signs, vehicles, pedestrians, etc. We categorize the elements that make up the traffic scene, and then define them in a collective way by using graph models.

The traffic scene elements are divided into two categories: background elements and foreground elements. Let represents the traffic scene elements of the ith image, denotes the background scene of the ith image, is the foreground scene of the ith image. The traffic scene structure is shown in Figure 2(a). We define the traffic scene of the i_{th} image as follows:

$$G_i = \{BS_i, FS_i\} \tag{1}$$

The background scene is composed of "sky", "pavement", "left wall", "right Wall" and "background wall". The background elements are represented as a set:

$$BS_i = \{SK_i, RS_i, LW_i, RW_i, BW_i\} \tag{2}$$

where SK_i, RS_i, LW_i, RW_i and BW_i denote "sky", "pavement", "left wall", "right wall" and "background wall" of the i_{th} image respectively.

The foreground scene is FS_i composed of traffic lights, traffic signs and moving objects:

$$FS_i = \{L_i, TS_i, MO_i\} \tag{3}$$

where L_i denotes the traffic light, TS_i is the traffic sign, and MO_i represents the moving object.

Then we define the set of traffic lights L_i, which is represented as follows:

$$L_i = \{TL_i, SL_i, WL_i\} \tag{4}$$

where TL_i, SL_i and WL_i denote "traffic light", "street light", and "warning light", respectively.

The set of traffic signs is defined as:

$$TS_i = \{WS_i, LL_i, ZS_i\} \tag{5}$$

where WS_i, LL_i and ZS_i denote the "warning sign", "lane line" and "zebra crossing", as shown in Fig. 2(c).

Finally we define the moving objects, which is defined as:

$$MO_i = \{VC_i, PS_i\} \tag{6}$$

where VC_i and PS_i denote the "vehicle" and the "pedestrian", as shown in Fig. 2(d).

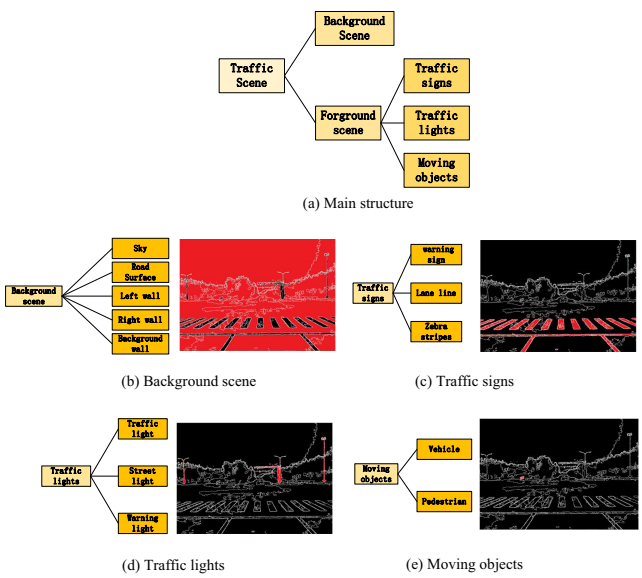

Fig. 2. Graph model of the traffic scenes. (a) Main structure. (b) Background scene. (c) Traffic signs. (d) Traffic lights. (e) Moving objects.

IV. Scene models construction based on scene stages

A. Scene stage description

Assuming that a person's viewpoint is a specified, then the perception space of the person can be represented by a cube. In front of the viewpoint is a wall perpendicularly to the ground, which is called "background". The background may be an outdoor mountain in real scene, or a wall in front of the room. Below the viewpoint, The area perpendicularly to the "background" is "pavement". The "pavement" can be the road surface or indoor floor. The two sides of the cube on the left and right sides of the viewpoint are called "left wall" and "right wall". The can be trees, buildings on both sides of the road, or the ordinary walls of a room. The top of the viewpoint is called "sky", which may be blue sky or a ceiling. Besides, there may be detailed objects in the real scene. We can describe these objects inside a cube that can roughly describe the scene layout, which is called scene stage. The 3D scene layout is characterized by color modeling.

Scene stage is a 3D geometric model of the scene. The scenes are divided into a set of finite image-level 3D geometry classes called "stages" to represent the general scene geometry. We use scene stage to determine the color used in different geometric regions. The target image can be divided into several stages, and a rough 3D geometric model is then constructed. In this paper, we focus on the analysis of traffic scenes. Based on the segmentation of the image, a 3D geometric model can be constructed.

B. Establishement of traffic scene stages

Traffic roads are used to provide access for trackless vehicles and pedestrians. The traffic roads can be classified into highways, urban roads, and rural roads. In the traffic road, we can easily observe that different road locations present different real scenes. The traffic scenes are basically composed of a stable geometric environment. After comparing and analyzing the images of various traffic roads, we observe that different road types, e.g. highway, urban road and country road, can be defined according to the road structure. These structures include straight road, turn left, turn right, intersection, tunnel, etc. We try to construct a scene wireframe model based on the structural laws of 2D images of traffic scenes. In order to construct and simulate the traffic scenes, we need to divide the road regions according to the road image content to determine the type of scene stages. Furthermore, we construct traffic scene wireframe model. The 3D traffic scenes are then constructed based on the wireframe model. When the road image is divided into different regions, we can partition the road into "background wall", "left wall", "right wall", "ground" and "sky". The yellow, red, green, gray, and blue patches are applied to represent the "background", "left wall", "right wall", "ground" and "sky", respectively. The characteristics of road scene are described by scene frame model, as shown in Fig. 3.

The potential number of traditional semantic scene categories can be very large, and the specification of high-level semantics from images is still difficult and unreliable [2]. Compared to the semantic scene classification, the traffic scene wireframe model can express the road's depth information more intuitively, and provide the basic model for the simulation of 3D traffic scenes.

C. 3D traffic scenes construction

Scene stage describes a traffic scene wireframe model based on the analysis of road images, which represents the scene layout for 3D traffic scene modeling. Traffic scene images are classified into simple and complex types based on the scene layout. Simple road images consist of basic elements of traffic scenes, while complex road images consist of both basic and foreground elements.

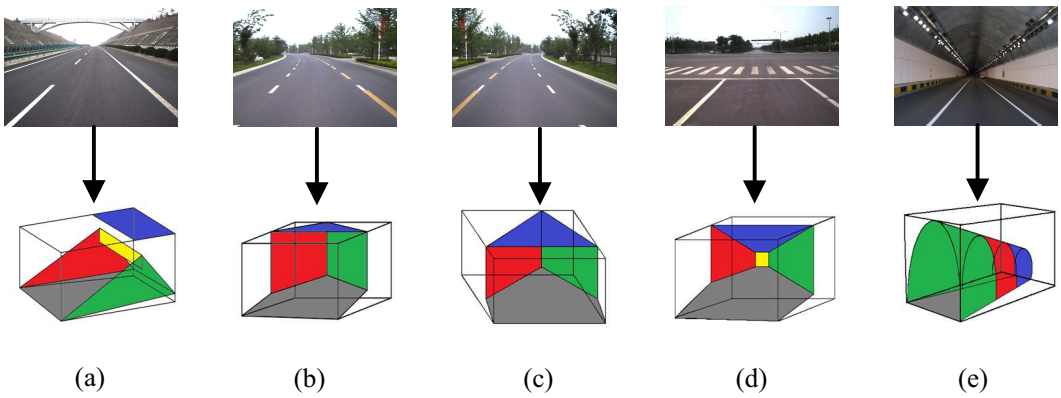

Fig. 3. Scene Stages of Traffic Scene. (a) Through Lane. (b) Left-turn Lane. (c) Right-turn Lane. (d) Intersection. (e) Tunnel.

We determine the control points of road boundaries based on the traffic scene wireframe model. The control points of road boundaries are distributed on both sides of the road [4]. In the scene stage, two farthest control points are taken at the far end of the road surface, while two closest control points are specified by the near end of the road surface. The remaining control points are distributed between the farthest and the nearest control points, as shown in Fig. 4(a). We then construct a 3D corridor road scene model based on the control points (the principle is shown in Fig. 4(b)). The texture mapping of road images from 2D to 3D is shown in Fig. 5.

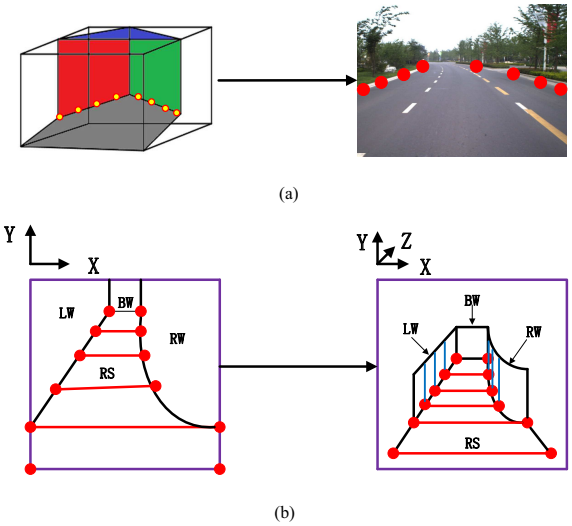

Fig. 4. Generation of control points. (a) Labeling Border Control Points. (b) Scene Modeling Principle.

The reconstruction of complex road images usually include bridges, vehicles and other traffic elements, as shown in Fig. 5. We refer these elements as foreground objects, which need to be processed by three steps: (1) extraction of foreground image, (2) modeling of basic scene, and (3) restoring of foreground image. The construction examples of scene models is shown in Fig. 6.

Our method is also applicable to generate cartoon scene models. We use opencv for cartoonization by four steps: (1) Apply bilateral filters to reduce the image color. (2) Convert the color image to grayscale and use a median filter to reduce image noise. (3) Use adaptive threshold to create contours. (4) Superimpose the colorful image in step 1 with the contours from step 3. After the generation of cartoon images, we construct traffic scene models with cartoon style, as shown in Fig 6 (c).

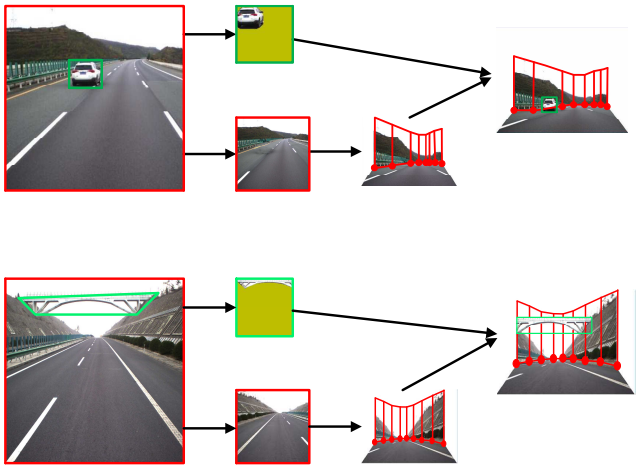

Fig. 5. Construction of complex models.

V. Experiments and Applications

The platform for our experiments is a computer with Intel core i5 processor @3.33 GHz and 8.00GB RAM. The experiments for road region detection are implemented with MATLAB R2014a, while the scene construction experiments are based on OpenGL Toolbox. Both of them are mainly based on the TSD-max Dataset [15], which is constructed by the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University.

A. Experiments and comparisons

Google Street View Map and Microsoft Street Slide provide users with scene knowledge by restoring 3D scenes. We

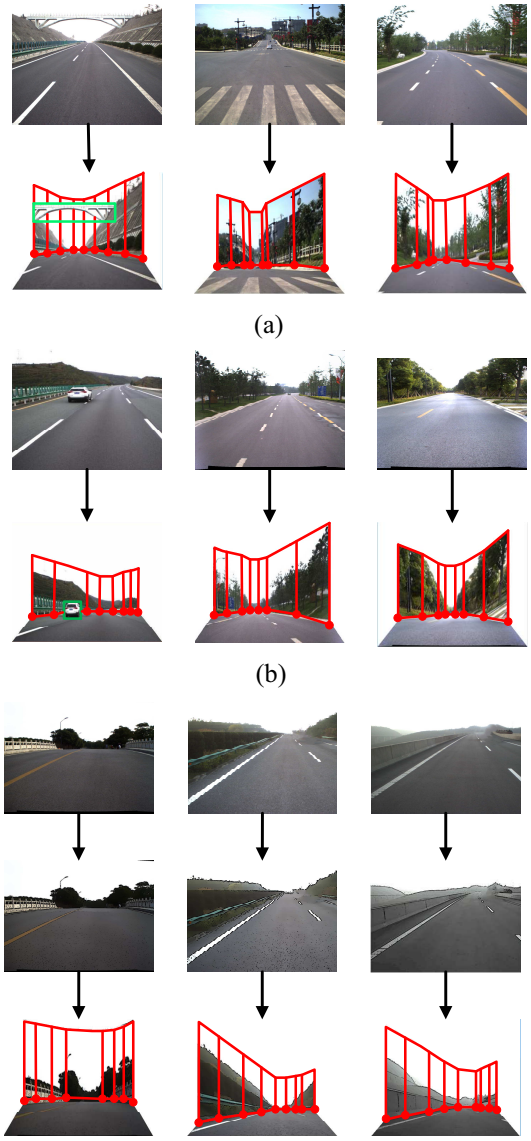

Fig. 6. Construction of scene models. (a) and (b) Traditional scene models. (c) Cartoon scene models.

perform the 3D traffic scene simulation based on the 3D corridor road scene model. The viewpoint is freely changed in the program, which is controlled by the commands of forward, backward, upward, and downward. According to different move types of the viewpoint, the simulation of traffic scenes is classified into two modes: (1) Roaming mode. The new viewpoint images are generated according to the viewpoint movement. (2) Bird view mode. The viewpoint is static, and the virtual vehicles and obstacles can be supplemented into the scene. The generation of new viewpoint images is shown in Fig. 7. The black grid represents the basic scene model, while the green grid represents the foreground vehicle model. The users can change the view angle based on the road scene model. The new viewpoint images are then generated, as shown in Fig. 7(b).

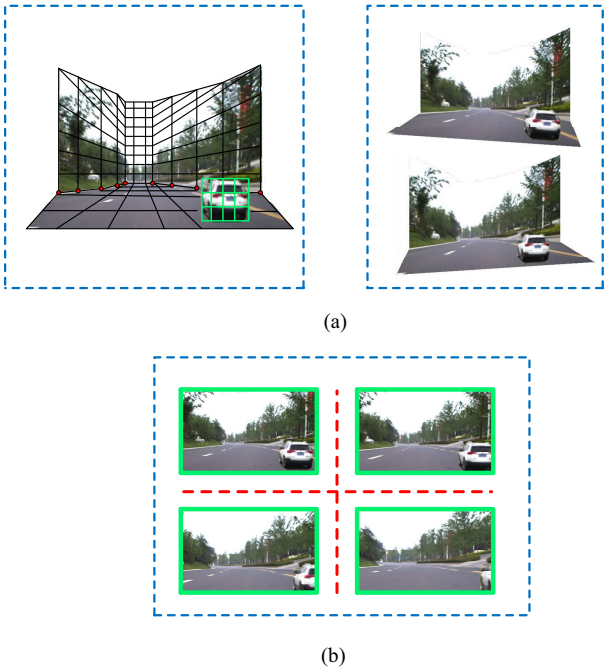

Fig. 7. Generation of new viewpoint images. (a) The control of view angles.The black mesh denotes the background scene, while the green mesh denotes the foreground vehicle. (b) The new viewpoint images.

We compare the proposed system with Google Street View and Microsoft Street Slide, as shown in Table 1. GSV denotes Google Street View, while MSS denotes Microsoft Street Slide. Four functionalities are compared: (1) Independent Foreground. (2) Speed Control. (3) Free Viewpoint. (4) Cartoon Rendering. As the comparison results show, the foreground objects are modeled independently in our system. The speed and viewpoint can be controlled by the users. Moreover, the cartoon-style scene models are constructed by our method.

TABLE I
SCENE FUNCTIONALITIES COMPARISON

	Foreground Modeling	Free Viewpoint	Speed Control	Cartoon Rendering
GSV	√	×	×	×
MSS	√	×	×	×
Ours	√	√	√	√

Furthermore, we evaluate the proposed method with the methods of Make3D [13] and Photo Pop-up[14] based on TSD-max Dataset, as shown in Fig. 9. Compared to Make3D and Photo Pop-up, the scene construction based on our method is more realistic. Besides, the scene structure of our method is more stable.

B. Applications

3D traffic scene modeling provides a testing framework for the off-line evaluation of unmanned vehicles. The simulation of traffic scenes make it possible to test the scenes that are dangerous in the real environment. The off-line test is

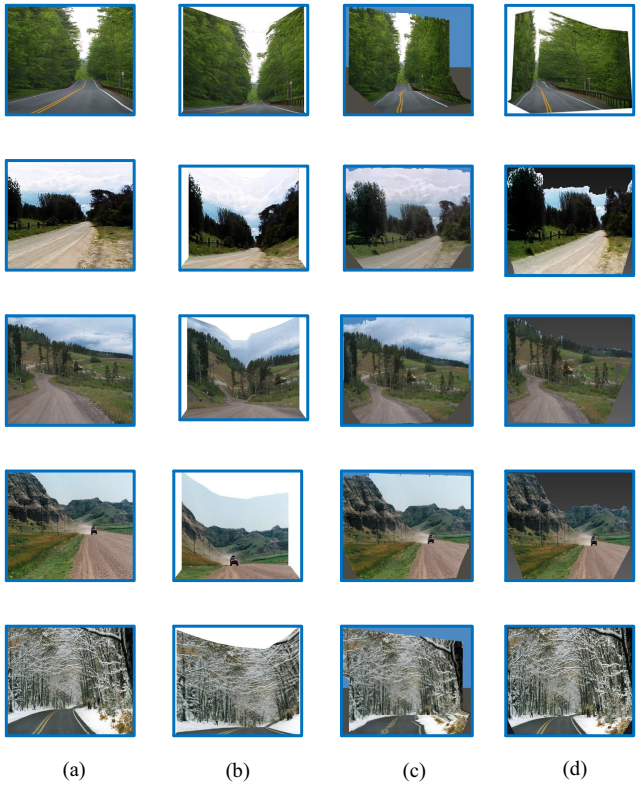

Fig. 8. Comparison of 3D Traffic Scene Construction methods. (a) Original images. (b) Results based on our method. (c) Results based on Make3D. (d)Results based on Photo Pop-up.

safer, more reliable and saves time and energy. Moreover, a specific test scene can be used repeatedly, and thus brings great convenience to the research of unmanned vehicles.

We propose five metric designed for the offline test, including pedestrian recognition, collision avoiding, traffic signal recognition, pavement identification, and the recognition of fog. Four levels of performances are defined to evaluate the vehicle behavior: (L0, L1, L2, L3), as shown in Table II.

The complexity of the road scene is generally divided into three categories denoted by (R1, R2, R3): road conditions, special areas and special weather. Based on the these categories, we can reconstruct various scenes and easily perform off-line test for unmanned vehicles. The three categories of modeling are as follows:

- R1: Different road conditions can be divided into: rural road, urban road, highway and tunnels. The different road conditions involve specific road width, the number of obstacles, traffic signs and traffic lights.
- R2: Special areas include campus areas, hospital areas, crowded streets, etc. Vehicles in the special areas need to react quickly and perform different operations.
- R3: Special weather is divided into rainy, snowy and foggy. The degree of the special weather is vary from each other. Therefore, it is necessary to accurately identify them and respond for the vehicles.

Finally, we can evaluate the performance of unmanned vehicles by the combination of the scene complexity and the evaluation metrics.

TABLE II
EVALUATION METRICS OF UNMANNED VEHICLES

	L0	**L1**	**L2**	**L3**
Pedestrian detection	Partly detect	Most detect	Totally detect	Perfectly detect
collision avoid	Completely collide	Always collide	Slghtly collide	Perfect avoid
Traffic signal recognition	Partly identify	Most identify	Totally identify	Perfect identify
Pavement identify	Not at all	misidentify	recognize	Perfect recognize
Rain and fog identify	Not at all	Misidentify	Recognize	Perfect recocgnize

VI. Conclusion and Future Works

In this paper, we present a method for 3D traffic scene construction and simulation based on scene stages. Firstly, the road images are analyzed by semantic labels, and the elements of the traffic scenes are defined by the graph model. The scene stages are defined based on analysis results. The scene stages illustrate the semantic analysis and the scene layout explicitly, which provide an important reference for the reconstruction of 3D traffic scenes. The proposed method has a wide range of application prospects.

In the future, we will apply depth maps to 3D traffic scenes construction and enrich the semantic details of the scene. More applications for the simulation and evaluation of unmanned vehicles will be explored and developed.

References

[1] Z. Lou, T. Gevers, N. Hu. "Extracting 3D Layout From a Single Image Using Global Image Structures," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3098-3108, 2015.
[2] Nedovic V, Smeulders AW, Redert A. "Stages as models of scene geometry, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1673-1687, 2010.
[3] Brostow J, Fauqueur J, Cipolla R. "Semantic object classes in video: A high-definition ground truth database, " *Pattern Recognition Letters*, vol. 30, no. 2, pp. 87-97, 2009.
[4] Li Y, Liu Y, Su Y, et al. "Three-Dimensional Traffic Scenes Simulation From Road Image Sequences, " *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1121-1134, 2016.
[5] Hoiem D, Efros A, Hebert M. "Recovering Surface Layout from an Image, " *International Journal of Computer Vision*, 2007: 151-172.
[6] Hedau V, Hoiem D, Forsyth D. "Recovering the spatial layout of cluttered rooms, " *IEEE International Conference on Computer Vision*, 2010: 1849-1856.
[7] http://www.map.google.com
[8] D. Anguelov, C. Dulong, D. Filip, et al., "Google street view: Capturing the world at street lebel," *Computer*, vol. 42, no. 6, pp. 32-38, 2010.
[9] J. Kopf, B. Chen, R. Szeliski, et al., "Street slide: Browsing street level imagery," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 102-106, 2010.
[10] L. Li, D. Wen, N. Zheng, et al., "Cognitive cars: A new frontier for ADAS research," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 395-407, 2012.
[11] http://www.cs.uiuc.edu/homes/dhoiem/projects/data.html
[12] http://make3d.stanford.edu/
[13] http://trafficdata.xjtu.edu.cn/index.do