Where Does the Driver Look? Top-Down-Based Saliency Detection in a Traffic Driving Environment

Tao Deng, Kaifu Yang, Yongjie Li, *Member, IEEE*, and Hongmei Yan

***Abstract*—A traffic driving environment is a complex and dynamically changing scene. When driving, drivers always allocate their attention to the most important and salient areas or targets. Traffic saliency detection, which computes the salient and prior areas or targets in a specific driving environment, is an indispensable part of intelligent transportation systems and could be useful in supporting autonomous driving, traffic sign detection, driving training, car collision warning, and other tasks. Recently, advances in visual attention models have provided substantial progress in describing eye movements over simple stimuli and tasks such as free viewing or visual search. However, to date, there exists no computational framework that can accurately mimic a driver's gaze behavior and saliency detection in a complex traffic driving environment. In this paper, we analyzed the eye-tracking data of 40 subjects consisted of nondrivers and experienced drivers when viewing 100 traffic images. We found that a driver's attention was mostly concentrated on the end of the road in front of the vehicle. We proposed that the vanishing point of the road can be regarded as valuable top-down guidance in a traffic saliency detection model. Subsequently, we build a framework of a classic bottom-up and top-down combined traffic saliency detection model. The results show that our proposed vanishing-point-based top-down model can effectively simulate a driver's attention areas in a driving environment.**

***Index Terms*—Traffic environment, bottom-up, top-down, visual attention, saliency detection.**

I. Introduction

A traffic driving environment is a complex and tridimensional scene with multiple information sources, which changes dynamically and requires instant processing, especially in an urban road. Suppose that you are driving a car to visit a friend; you navigate to the desired destination while paying attention to different types of objects in the environment (e.g., roads, traffic signs, cars, pedestrian, and the street) and obeying traffic laws (e.g., speed limit and stop signs). Do you know where you should and might look at in different traffic situations? What salient areas should you mainly concentrate on when you are executing these task-based behaviors? Of course, you manage these competing tasks by selectively fixating your eyes on the most important and salient areas or targets instantaneously and effortlessly according to your driving experience and the immediate demands. However, how does an autonomous car detect the traffic saliency to achieve safe driving?

Saliency detection is one of the fundamental function in human vision, and it is also one of the important tasks in computer vision applications, such as object detection and recognition [1]–[3], scene understanding [4], robot navigations [5], image processing of the unmanned intelligent vehicle, driving simulators (e.g., driver assistant systems), and other visual related control systems. As human visual attention is naturally attracted towards visually salient stimuli, understanding the mechanisms of visual attention and then modeling the selective processing of visual scenes are the most ideal and promising solution for saliency detection algorithm.

Although the ability of the human visual system to detect visual saliency is exceedingly fast and reliable, computational modeling of this basic intelligent behavior still remains a large challenge [6], [7] for the following reasons. 1) Human and biological vision can interrogate complex, noisy, dynamic environments to accomplish saliency detection tasks automatically. However, how are these tasks performed so effortlessly and reliably? What type of control structure is robust when facing the complicated and varying nature of the visual world? [8] Over the past few decades, a considerable amount of experimental research has been conducted, and some theories have been proposed, such as Broadbent's Filter Model of Attention [9] and the Feature Integration Theory of Treisman and Gelade [10]. However, the neural mechanisms of this selective attention and the parallel processing of our brain are still largely unclear. 2) It is widely agreed that visual attention operates in both bottom-up and top-down modes. Therefore, there are two factors that influence visual saliency. One factor is the bottom-up, task-independent factor, which is driven by the low-level attributes of input images, such as color, intensity, and orientation. The other is the top-down, task-dependent factor, which is driven by tasks, goals and experiences and other contributors. The famous example of top-down attentional guidance was presented by Yarbus in 1967 [11]. He illustrated that human eye movements largely depend on the specific tasks in the experiment. In other words, visual saliency can be either object- and feature-based or task- and experience-based.

Generally speaking, bottom-up attentional modeling that is based on the intrinsic image features is relatively simple and easy. In fact, starting from the Feature Integration Theory of

Manuscript received March 31, 2015; revised July 18, 2015; accepted September 21, 2015. Date of publication March 25, 2016; date of current version June 24, 2016. This work was supported in part by the 973 Project under Grant 2013CB329401; by the Natural Science Foundation of China under Grant 91120013, Grant 91420105, Grant 61375115, Grant 31300912, and Grant 61573080; by the Fundamental Research Funds for the Central Universities under Grant ZYGX2013J098; and by the 863 Project under Grant 2015AA020505 and the 111 Project (#B#12027). The Associate Editor for this paper was L. Li.

The authors are with the Key Laboratory for Neuroinformation of the Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: tinydao@163.com; yang_kf@163.com; liyj@uestc.edu.cn; hmyan@uestc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2535402

Treisman and Gelade [10] and the bottom-up attentional model of Koch and Ullman [12], a series of ever-refined algorithms have been designed to predict where subjects will fixate in synthetic or natural scenes [6], [13]–[18]. Typically, Itti *et al.* [13], [14] proposed an amazing bottom-up-based saliency detection model (usually called the Itti Model), in which multiple low-level visual features, such as the intensity, color, orientation, and texture, were extracted from the image at multiple scales. They computed saliency maps of each feature and then normalized and combined them in a linear or non-linear fashion into a master saliency map that represents the saliency of each pixel. Finally, the winner-take-all and inhibition of return operations were adopted to identify every significant area. These models have been able to account for an increasing fraction of human eye fixations during free viewing.

In addition, other saliency detection algorithms have been developed in recent years. For example, Harel *et al.* [18] proposed another bottom-up visual saliency model called Graph-Based Visual Saliency (GBVS). These authors computed a saliency map by defining Markov chains over various image maps and then normalizing them in a way that highlights conspicuity and admits combination with other maps. This model is simple and biologically plausible. Hou and Zhang [6], [19] proposed a simple and fast algorithm called the spectrum residual algorithm (SR), which was based on the Fourier Transform. They proposed that the spectrum residual corresponds to the image saliency. Bruce and Neil proposed a bottom-up model called Attention based on Information Maximization (AIM) [20], and Zhang proposed a Saliency Using Natural statistics (SUN) model [21]. All of these bottom-up saliency models have been very successful at predicting human visual saliency in free viewing and visual search. However, they lack top-down control. Therefore, they have limits in explaining the specific or task-relevant saliency in everyday tasks, such as driving [15], [22], [23].

Traffic saliency detection, which computes the salient and prior area or targets in a specific driving environment, is an indispensable part of intelligent transportation systems (ITS). However, as described previously, driving is a dynamic, task-oriented behavior. It is possible that the real traffic saliency addressed by human fixations might be completely different from the salient maps computed by traditional saliency algorithms or models. Fig. 1 shows examples in which the classical bottom-up-based saliency maps computed by the GBVS and Itti saliency models do not match the driver's actual attention or fixational areas in simulating driving tasks and even do not match those in free-viewing tasks (actual human saliency maps are reflected by human fixational eye movements from eye-tracking recording). Although we show only the saliency maps obtained by the GBVS and Itti models here, other typical models like AIM, SR and SUN models were also considered, and none of these models can accurately estimate the actual human attentional areas in driving conditions. In other words, these models can explain only a small portion of the human fixations in a real traffic driving environment.

From above, top-down attention should be considered to build a better model of the visual behavior in a driving task. However, a top-down attentional model is much more complicated because it is task-dependent. Different tasks may require different algorithms, and there are often several factors (e.g., task difficulty, environment, distractors, subject's experience and actions) that must be considered in different tasks, especially in the context of a long dynamical temporally extended task.

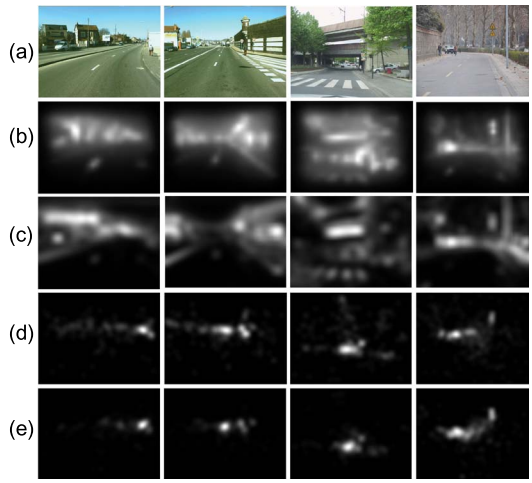


Fig. 1. Existing bottom-up saliency models are difficult to accurately estimate human salient areas in a traffic environment. (a) Original images, (b) saliency maps computed by the GBVS model, (c) saliency maps computed by the Itti model, (d) saliency maps viewed by a non-driver's free viewing, and (e) saliency maps viewed by a driver's simulating driving.

Researchers have attempted some top-down attentional modeling over the past two decades. Rimey and Brown [24] attempted to model the top-down attention using Bayes nets and decision theory for scene domains and tasks. In 2001, Itti and Koch proposed the idea of top-down influence to better estimate the saliency in specific tasks [15]. They considered that there was a link between visual attention and eye movement. Thus, it is necessary to combine eye movement with a computational model to study the human visual system. In recent years, some top-down saliency models have been proposed with learning methods. Judd *et al.* [17] considered top-down information in their research by designing an eye-tracking experiment to collect eye-tracking data in their dataset and build a saliency learning model. This model produces a saliency map by analyzing the low-, mid- and high-level features of the input image and then combining them after training the features for every pixel of the image. Zhao *et al.* [25] also proposed a saliency model that was based on a learning method. They computed the weights of the features, such as the color, intensity, orientation and face, by statistically analyzing the eye-movement data. In their work, the weight of the face is higher than that of other features. The performances of the above two top-down models show great improvement. In 2005, Navalpakkam and Itti [1] proposed guidelines for modeling the influence of a task on attention in natural scenes. Subsequently, Peters and Itti [26] learned a mapping from the global context of scene-to-eye fixations using the data of subjects playing contemporary video games, and they evaluated the relative importance of bottom-up and top-down factors at the time of an event [27]. They used the multiplication of bottom-up saliency and their top-down fixation prediction. In driving situation, Lim *et al.* [28] proposed a queuing network-based computational model to

simulate driving performance in a pedestrian-detection task. They found that driver's different eye-movement strategies generate different eye-movement behaviors, then they used driver's eye movement strategies as top-down reinforcement learning process in pedestrian-detection, indicating the potential strength of a cognitive based model to investigate the effectiveness of an ITS.

Most of the above top-down saliency models make use of the eye movement information and can be applied to some simple tasks such as reading and visual searching. However, how to effectively integrate the bottom-up salient and top-down task-driven control together in a complex, interactive, and temporally extended task remains a substantial problem. Recently, Itti's group described new task-dependent approaches for modeling top-down overt visual attention based on graphical models for probabilistic inference and reasoning [7]. They presented a general framework for interpreting human eye-movement behavior that explicitly represents the demands of many different tasks, perceptual uncertainty, and time by an example of video games. In their papers, bottom-up and top-down models were integrated into a state-of-the-art visual attentional model. However, this framework had not been used in a traffic environment. Recently, some saliency models were applied in traffic sign detection [29]–[31]. There is still a lack of experimental research and saliency models to predict a driver's real attention and gaze areas during driving.

As described previously, driving is a specific top-down attention-dependent task. Visual attention and eye movements are closely related, although this link is not perfect since covert visual attention can occur without eye movements [32]. Nevertheless, eye movements and visual attention are linked in most instances [15], [33]. Drivers' eye movements often provide a clear window into the minds of drivers in a way that sometimes allows inference of how drivers solve competing objectives and how they maintain priority for the most important visual information in a selection mechanism. Previous behavioral studies of driving analyzed drivers' daytime eye movements and found that drivers looked straight ahead at the road 59 percent of the time, to the right side of the road 15 percent of the time, and to the left side of the road 25 percent of the time [34]. Both Paul Green's work [35] and Panos Konstantopoulos's work [36] or other related researches do lots of work on driver's visual attention with behavioral experiments, lots of meaningful results such as driving experience, areas of interest and eye movements have been revealed in this field. Our behavioral experimental results are also consistent with previous studies. Our results from eye-tracking experiments also showed that the viewers' gazes mostly concentrate on the vanishing point of the road [37]. Similar with the idea that driver's eye movement strategies can be used as top-down reinforcement learning process in pedestrian-detection model in Lim's study [28], the driver's gazes strategy is also a very important aspect of top-down control when driving. Therefore, the focus of this paper is to make use of the top-down attentional mechanism to build a top-down traffic saliency model. An algorithm proposed by Kong *et al.* [38], [39] is mainly adopted for vanishing point detection in this paper. The results show that the vanishing point-based top-down traffic saliency model has an amazing improvement on the traffic saliency detection compared to the classic bottom-up models. This work is based on our previous work reported in a conference proceeding [40], which is substantially extended here and more performance analysis.

The remainder of this paper is organized as follows. In Section II, we describe the details of behavioral psychophysics and data analysis of eye movement. In Section III, we present the framework of our top-down-based saliency detection model and algorithm. In Section IV, we evaluate the performance of the proposed model with the actual-movement data. Finally, we discuss our model and draw conclusions in Section V.

II. Psychophysics and Eye-Movement Analysis

The arbitration process of visual perception incorporates and uncertainty priority in two modes of attentional processing, bottom-up and top-down attention. In most real-life situations, the responses of the nervous system to a sensory input depend on both bottom-up influences as driven by the sensory stimulus and top-down influences that are shaped by extra-retinal factors such as learning, past experience, and the current state and goal of the task [41]. In the traffic driving task, a driver's traffic saliency detection is also influenced by both the bottom-up and top-down modes. To better understand the regions of visual interest in a driving environment and to make full use of the top-down attentional mechanism to build up an effective traffic saliency detection model, we have collected and compared two groups of human traffic saliency databases with eye-movement tracking in behavioral experiments that involve both attentional mechanisms.

A. Stimuli and Subjects

A total of 100 traffic driving images were collected, all of which were from urban roads (Fig. 1). A total of 40 subjects (18 male and 22 female) aged 21–45 years old (average age 28) were enrolled to participate in the experiment. The subjects were divided into two groups, and each group comprises 20 subjects. One group has no driving experience (Group I), and the other group comprises drivers who have at least two years of driving experience (Group II). All subjects were blind with respect to the purpose of the experiments. Group I was asked to view the images freely, and group II was told to view the images while assuming that they were driving a car. Thus, the two groups were driven by two different attentional mechanisms when executing the tasks: group I was driven mainly by the bottom-up attentional mode, and group II was driven by both the bottom-up and top-down attentional modes. Each image was presented at full resolution for 10 seconds, with a 20-second separation for rest between images using a gray screen. The participants' eye movements were recorded with an infrared eye tracker (Eyelink2000, SR Research Ltd.). The participants' head movements were restricted by a forehead and chin rest. The pupil of the left eye was tracked at a sample rate of 1000 Hz and a spatial resolution of approximately 0.1°. The experimental paradigms were approved by the Ethics and Human Participants in Research Committee at the University of Electronic Sciences and Technology of China in Chengdu, China.

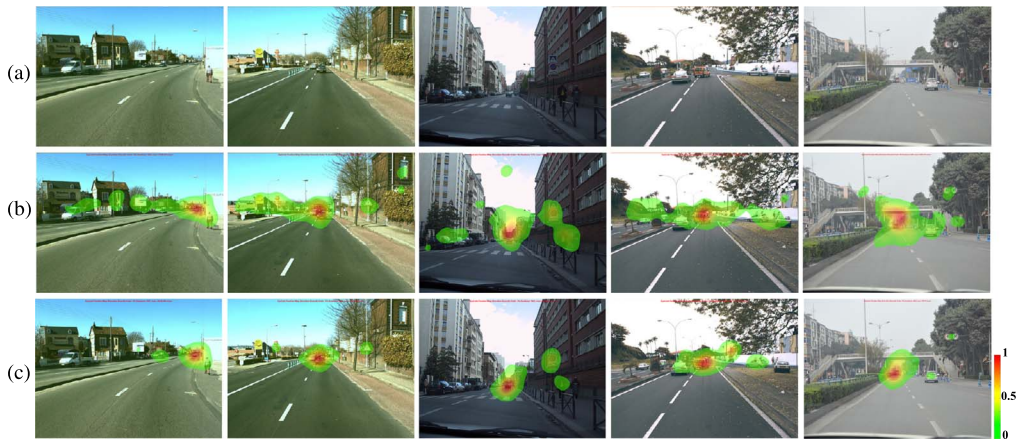


Fig. 2. Saliency maps driven by different selective attention in the experiment. (a) Original images. (b) Saliency maps freely viewed by subjects without driving experience. (c) Saliency maps of drivers with 2 years of driving experience when simulating driving.

B. Eye-Movement Analysis

The subjects' eye fixations and fixation durations were recorded to construct the human saliency map. To obtain a continuous saliency map of an image, we convolve a Gaussian filter across the user's fixation locations based on the fixation time. Examples of the average saliency maps of the two groups of viewers are shown in Fig. 2, in which the red and yellow areas that overlap on the image indicate that the areas are more observably fixated; the green areas indicate the locations that are comparatively observed, and the remaining areas have few fixations.

C. Saliency Difference Between Bottom-Up and Top-Down Attention

We compared the saliency maps driven by the bottom-up and top-down attentional mechanisms when the subjects viewed these scenes. We found that there were some significant differences between the two groups of saliency maps, although they shared some common areas.

First, the attentional areas of the non-experienced viewers are sparser and broader, but the areas of the driving-experienced group are more intensive and narrow (see examples in Fig. 2). Second, short-time fixations are scattered much more randomly on the images for the non-experienced group, whereas there are more long-time fixations concentrated on certain areas for the driver group. For example, for a non-driver (Fig. 3(b)), you could find that his/her gaze might saccade from the road ahead to a traffic light, then to a car nearby, then to an advertisement texture, then to the traffic light, and so on. The eye-movement paths are normally disordered. In comparison, for a driver (Fig. 3(c)), you could find that his/her gaze paths are more regular. The gazes might focus on the road ahead for a long time, then saccade to the traffic light and then back to the road quickly again. Third, the bottom-up saliency maps are varied among the different traffic environments. The most salient area of the bottom-up saliency maps might overlap on the end of the road in front of the vehicle or at traffic signs, traffic lights, pedestrian, advertisement textures and other odd targets in the environment, whereas the top-down saliency maps are of drivers who focus mostly on the end of the road in front of the vehicle.

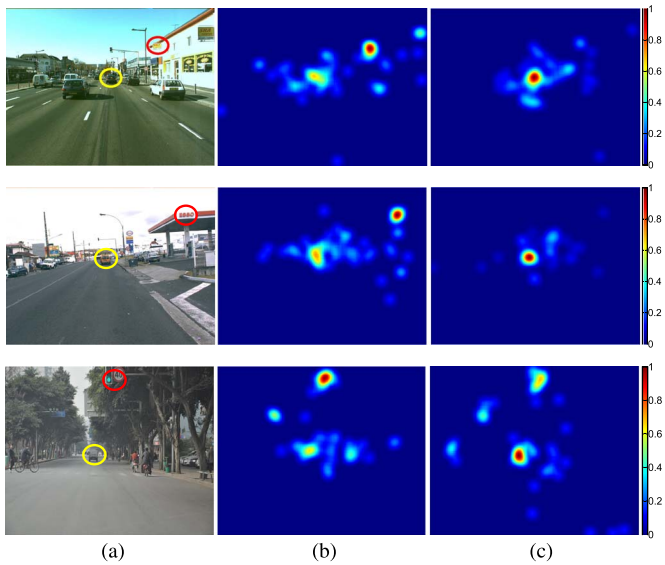


Fig. 3. Comparison of saliency maps driven by bottom-up and top-down selective attention within the first 2 seconds. Column (a) shows the sample images in the traffic road dataset. The areas marked in red circles correspond to the most attractive and salient areas as driven by the bottom-up attentional mechanism. The areas marked in yellow circles correspond to the most interesting and important areas as driven by the top-down attentional mechanism. Column (b) shows the eye-tracking saliency maps of subjects without driving experience under the bottom-up attention mechanism. Column (c) shows the eye-movement saliency maps of drivers under the top-down attentional mechanism.

Fig. 3 illustrates some examples of the difference between the bottom-up and top-down saliency areas during the first 2 seconds of viewing. In Fig. 3, the areas marked in red circles in the original images correspond to the most attractive and salient areas as driven by the bottom-up attentional mechanism, where the physical features of the images are usually salient, such as salient color or orientation of a traffic light, textures, and so on. However, the areas marked in yellow circles correspond to the most interesting and important areas driven by the top-down attentional mechanism, which always indicate the end of the roads in front of the vehicle. However, from Fig. 3, the

human salient areas reflect a combination of bottom-up and top-down attentional mechanisms in a traffic environment. Therefore, both bottom-up influence and top-down control should be considered when modeling the traffic saliency detection.

D. Top-Down Information: Vanishing Point

The results of the experiment mentioned above show that the subjects, especially the drivers, focus most of their attention on the end of the road in front of the vehicle in a traffic environment (Fig. 2), although there are some differences between the eye movements of the two groups. This result is consistent with those of previous behavioral research about driving by Higgins *et al.* [34] and Underwood *et al.* [42]. We propose the fact that a driver's attention mostly focuses on the end of the road in front of the vehicle and might be endogenous top-down control or guidance in a traffic environment. Kong *et al.* proposed that there exists an important point, which is called the Vanishing Point (VP), in the road in front of the vehicle, and they propose a VP detection algorithm [42]. We find that the most salient areas that are of interest to both groups could overlap with the vanishing points of the roads in most cases, regardless of whether the roads are straight or curved roads. Thus, the VPs could be very important top-down information in a traffic saliency model.

III. Top-Down-Based Saliency Detection Model

According to the behavioral experimental results obtained above, a top-down-based saliency detection model is built in this section. In the following, the vanishing point detection algorithm alone is introduced first, and then, acting as a kind of top-down control, the detected VP is combined with bottom-up information. The proposed framework of a top-down-based traffic saliency model is graphically shown in Fig. 6, which is composed of a classical bottom-up saliency model and a top-down constraint.

A. Vanishing Point Feature and Detection Algorithm

The vanishing point (VP) feature of the road furnishes important information in traffic driving research. Research on the VP has become increasingly developed and ready for use. For the detection of the VP in the road, several algorithms have been proposed [38], [43]–[47]. Current methods can be generalized into two categories: edge-based and texture-based. The edge-based VP detection methods are mainly aimed at detecting the road borders, and the VP is marked as the intersection of the border lines. The texture-based methods aim to estimate the texture orientation of the images. In contrast to the edge-based approaches, the texture-based methods can accurately detect VPs on both well-paved urban roads and unstructured rural off-roads. In this paper, the texture-based VP detection method proposed by Kong *et al.* [38], [39], [47] is adopted to extract the VPs, and then, this information is regarded to act as top-down guidance in the traffic saliency detection model.

In the early work of Kong et al, they estimated the texture orientation with a Gabor-based method [38], [39]. Recently, they proposed a new generalized Laplacian of Gaussian (gLoG) filter [47] to estimate the texture orientation. The gLoG filter is applied to estimate the texture orientation at each pixel of an image. Then, the VP is detected based on the estimated texture orientations.

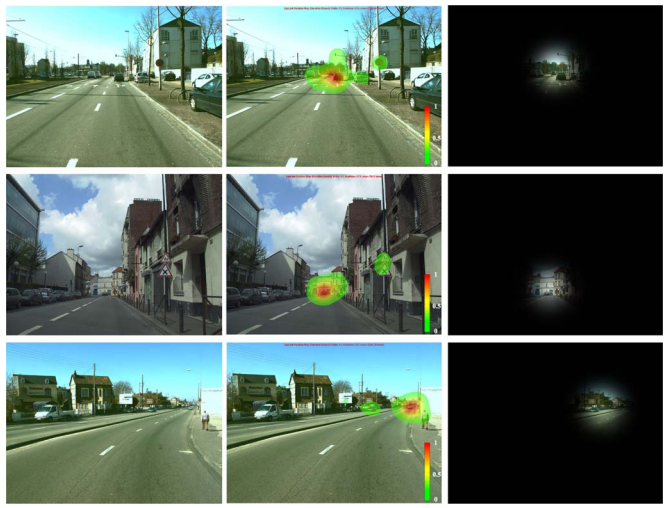


Fig. 4. Comparison of the detected vanishing points with the positions drivers fixate on most frequently. (a) Original images. (b) Saliency maps of drivers with 2 years driving experience. (c) The detected vanishing points by placing a 2-D Gaussian with $\sigma = 60$.

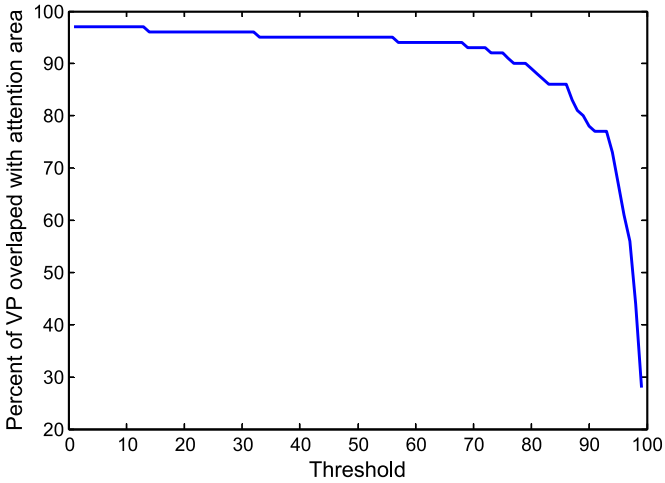


Fig. 5. The percent of the VP located in the subjects' attention area with different thresholds.

The experimental result shows that Kong's methods can estimate the VPs in our urban road image dataset. Nevertheless, not all VPs can be extracted successfully in our dataset. There are 4 (a total of 100) VPs are failed to estimate, which are affected by other distractors, such as the sky. Most of the VPs centralize on the end of the traffic road that is in front of the vehicle. However, some points are out of the central area, where the curved road exists.

We compared the vanishing points by placing a 2-D Gaussian with the eye-tracking saliency maps in the dataset qualitatively, and we found that the detected vanishing points overlap with the positions that drivers fixated on most frequently in most cases, as shown in Fig. 4. Therefore, we proposed that it is reasonable

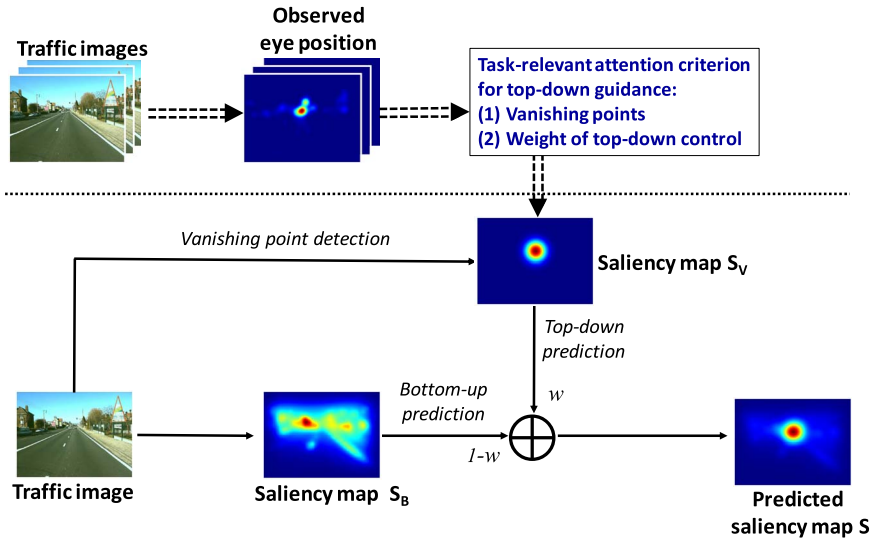


Fig. 6. The proposed framework of our method. The top part shows that we obtain the top-down control criterion by the drivers' eye position over a set of images based on task-relevant attention. The bottom part shows that for a given input image, we compute the bottom-up saliency map and the top-down saliency map separately based on a classical bottom-up model and top-down information (vanishing point), and then obtain the predicted eye positions by combining the bottom-up saliency map with top-down saliency map.

and feasible to regard the vanishing point of the road as a kind of top-down guidance in the traffic saliency detection model.

To analyze the relationship between the vanishing point and top-down visual attention quantitatively, we calculated the percent that the vanishing point overlaps with the attentional area at various thresholds (Fig. 5). We can directly find that the vanishing points overlap well with the attention area at most of the threshold values, with more than 90% of the VPs overlap with the attention area when the threshold is approximately 80%. This statistical figure also shows that the vanishing point is a kind of very important top-down information with respect to the visual attention during driving.

B. Bottom-Up and Top-Down Combined Saliency Model

As described in the introduction, many state-of-the-art bottom-up saliency models, including the GBVS, AIM, SR, SUN and Itti models, have been proposed. All of these saliency models show good performance in predicting human visual saliency during free-viewing of the natural scenes or object detection, but they could not be directly applied to predict special tasks or scenes such as a traffic driving environment (Fig. 1). In the following, we will show that combining the bottom-up saliency models with specific top-down control could effectively simulate the visual selective attentional mechanism in a driving task environment.

Based on the vanishing point information, we propose a computing framework of a top-down-based traffic saliency model (Fig. 6), which is composed of a classical bottom-up saliency model and a top-down constraint. The main methodology of the model is to find the top-down constraint on the input scene and then combine it with a classical bottom-up model in a linear fashion. Finally, the model constructs the saliency map $S(x, y)$ according to

$$S(x, y) = wS_V(x, y) + (1 - w)S_B(x, y) \quad (1)$$

Fig. 7. The computation of the percentage of the intersection of the eye-tracking saliency area and top-down saliency area in the total eye-tracking saliency area. Here, (a) shows the eye-tracking saliency area, (b) shows the top-down saliency area, and (c) shows the intersection of a and b.

where w is the weight, $S_V(x, y)$ represents the saliency map of the vanishing points convolved Gaussian filter, and $S_B(x, y)$ represents the saliency map of classical bottom-up saliency model.

C. Selection of the Weight

The weight w in Equation (1) is a key parameter of the proposed model, which implies the extent to which the top-down attention plays a role in the model. To select a proper weight, we have analyzed the relationship between the vanishing point placed in a 2-D Gaussian and the drivers' eye-tracking saliency map. The percentage of the intersecting area (Fig. 7(c)) between the two maps was calculated. Then, the average percentage was considered to be the weight. Here, because the vanishing points of the 4 images in the dataset were not detected successfully, the percentages of the intersecting areas of 96 traffic images were averaged.

The result of the aforementioned percentage is approximately 80%, which implies that in the drivers' attention area, the top-down saliency area could include 80% of the information that is mainly in the area around the vanishing point, and the bottom-up saliency area includes another 20% of the information, on features of images such as traffic signs. The result

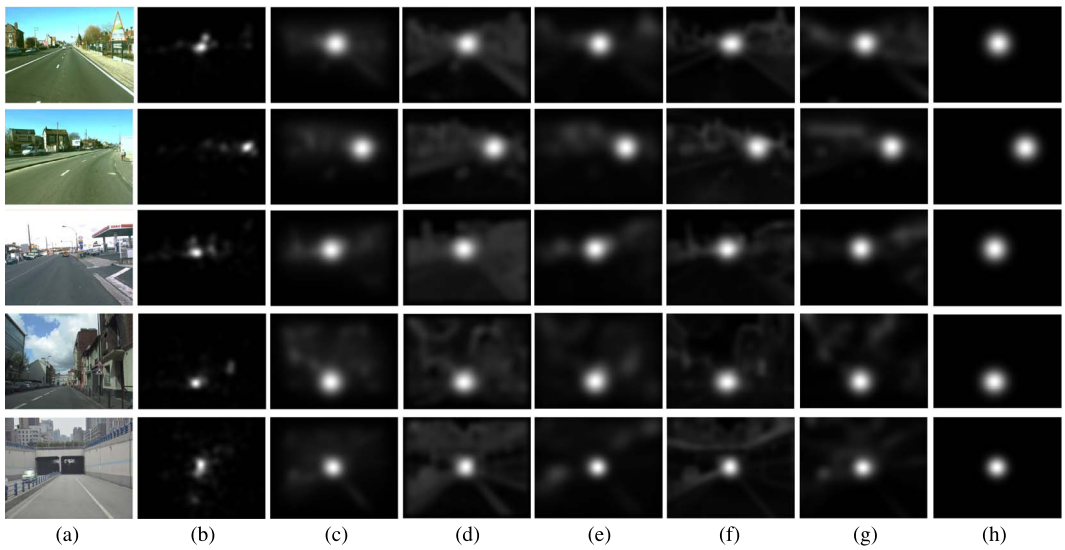


Fig. 8. Comparisons of the vanishing point based top-down saliency maps. (a) Original images. (b) Saliency maps of drivers based on top-down attention mechanism. (c)–(g) Bottom-up saliency models (GBVS, AIM, SR, SUN, Itti) combined with vanishing point guidance. (h) VP only maps.

also shows that most of the drivers' attention (approximately 80%) is focused on the vanishing point of the traffic road, and a small amount of attention (approximately 20%) is focused on the remainder of the scene. Besides, we have tested the different weights on the image dataset quantitatively, and we draw a conclusion that 0.8 is the most appropriate to be the weight w. The details are introduced and analyzed in the Section IV-C.

In the following, we combine the vanishing point information with five classical bottom-up saliency models, namely GBVS [18], SR [6], AIM [20], SUN [21] and Itti [13], [14]. We convolve the vanishing point with a Gaussian filter, and then, we combine the result with a classical bottom-up saliency map in a linear additive model. In the end, the final saliency map is obtained.

IV. Results

After integrating the classical saliency models (GBVS, SR, SUN, AIM and Itti) with the aforementioned VP based top-down information, we obtained the final saliency images. Here, we provided both qualitative as well as quantitative evaluation to assess the quality of the proposed approach benchmarking on our dataset collected in the psychophysical experiment. Four ways including the common Receiver Operating Characteristic (ROC) curve, Area Under the ROC Curve (AUC) values, the revised ROC curve, and Normalized Scan-path Saliency (NSS) scores were used to quantitatively evaluate our algorithms. Note that the fixation points collected from the non-driver subjects were used as the ground-truth to evaluate the bottom-up saliency models, while the fixation points obtained from the experienced drivers were used as the ground-truth to evaluate our final model, i.e., bottom-up saliency model joined with vanishing point information.

A. Qualitative Evaluation

Fig. 8 shows some examples of traffic saliency maps computed with the vanishing point based top-down models. The results show that the saliency maps computed by combined models not only include some other features of images such as traffic signs and lights but also include the most important top-down information with respect to the drivers. They can effectively simulate driver's attentional areas (column (b) in Fig. 8 in a driving environment. While the VP only maps (the last column in Fig. 8 are not accurate enough due to lacking of bottom-up information. In other words, the models we proposed realize the integration of the two visual attentional mechanisms: the bottom-up (feature-based attention) and top-down (task-dependent attention). The experimental results show that the algorithm performance is greatly improved when combining the bottom-up model with vanishing point information, which represents the top-down guidance.

B. Classical ROC Curve and AUC Value

Here, the model's saliency map is treated as a binary classifier on every pixel in the image. Pixels with larger saliency values than a threshold are classified as fixated while the rest of the pixels are classified as non-fixated. By varying the threshold, the ROC curve (Fig. 9) is drawn as the false positive rate vs. true positive rate, which can indicate how well the saliency map predicts actual human eye fixations [20].

Fig. 9 shows that the models combining the VP feature perform much better than the original models with bottom-up feature alone. From the black line representing the result of VP only model, we can find that the VP only model is better than the classical bottom-up saliency model, but worse than the VP combined models.

We also quantitatively analyze the performance with the AUC value after combining top-down information with each algorithm. Here, the ROC curve refers to aforementioned ROC method that consists of true positive rate and false positive rate [20], [48]. It can be seen from Table I that after adding top-down information to the saliency algorithms that use bottom-up information alone, the AUC values are significantly increased,

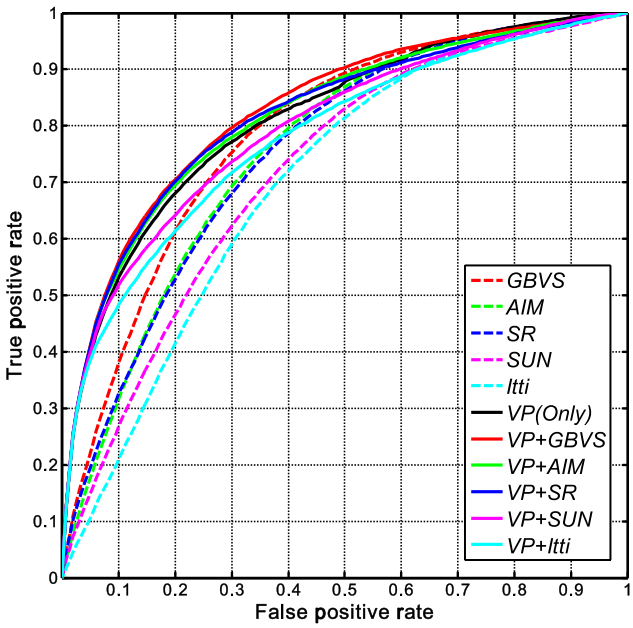


Fig. 9. The classical ROC curves of different algorithms. The solid lines show the ROCs of the algorithms combined with the VP information, the black line shows the ROC of VP only model, and the dashed lines show the ROCs of the classical saliency algorithms.

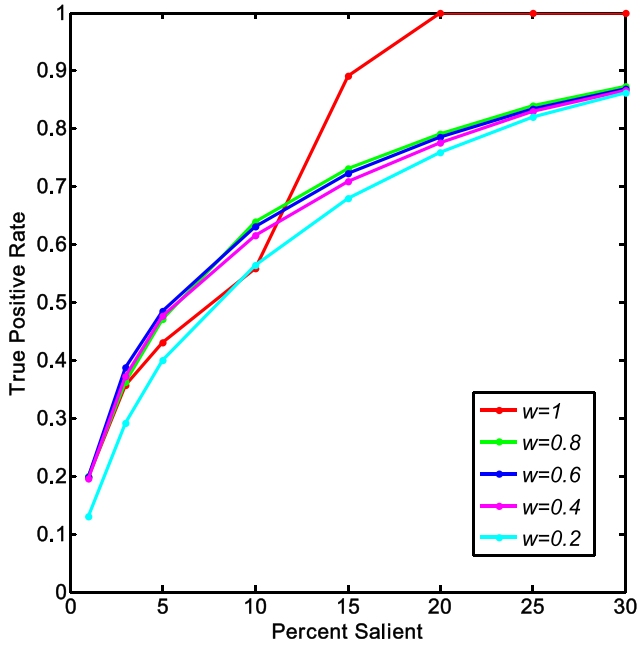


Fig. 10. The revised ROC curves of GBVS saliency model with different weights w in Equation (1).

TABLE I
A COMPARISON OF THE AUC VALUES OF EACH SALIENCY MODEL

AUC value	GBVS	AIM	SR	SUN	Itti	VP(only)
Without VP	0.7874	0.7581	0.7558	0.7210	0.7041	-
With VP	0.8286	0.8188	0.8181	0.7965	0.7809	0.8105
Improved	5.23%	8.01%	8.24%	10.47%	10.91%	-

which means that the new algorithm could simulate the human attentional mechanism better than previous algorithms. For example, the performance of Itti saliency model with VP top-down guidance is improved nearly 10.91% than that of without VP information. With VP only, we can see that the AUC value is higher than the classical bottom-up saliency models, but lower than some the saliency models with VP (e.g., GBVS+VP). Therefore, we conclude that the VP is a kind of effective top-down guidance in traffic saliency detection models, although bottom-up information is necessary for extracting task-independent salient regions.

C. Revised ROC Based on the Fixation Durations

Usually, the ground-truth saliency maps obtained by tracking eye-gaze consist of sparse binary fixation points (with value as 0 or 1), and lack of the information of fixation durations. In this paper, in order to evaluate the performance of the algorithms more accurately, we consider the different weights of the fixation points in the ground-truth saliency map based on the fixation durations. We first remove the extremely short ($<$ 100 ms) or long ($>$ 2000 ms) fixation points, because they may be unreliable in the visual attentional calculation. Then we linearly normalize the durations of each fixation and regard the value as the weight of each fixation point. Therefore, we obtain the ground-truth saliency maps with the value in the range of [0, 1], weighted by the fixation durations.

The classical computation of ROC is somewhat biased when benchmarking on the revised ground-truth weighted by the fixation durations. It can be explained as that fixation points located in the revised ground-truth are with soft values in [0, 1], but all background points are with the same value of 0 (i.e., without weights). This may result in the unbalance between true positive rate and false positive rate when computing the original ROC. Therefore, we revise the ROC referred as previous studies [17], [49], [50], where the ROC curve is conveyed with the true positive rate and percent salient. The salient regions are extracted by thresholding the saliency map at various percent salient level using the method described by [17], [49], [50]. Especially, the true positive rate is computed as the ratio of the summation of weighted fixation points within the salient regions to that of all weighted fixation points. Therefore, the revised ROC can reflect the power of fixational prediction at various percent salient level. With the increasing percentage of saliency, the true positive rates of all models will go up to 1, which means all the fixational points are located in the computed saliency maps. Thus, we can focus on the ROC below 30% saliency threshold for clearly demonstrating the comparison in this experiment, same as that in [17], [49], [50].

According to the quantitatively analysis, we further evaluate the value of parameter w in Equation (1). Fig. 10 shows the revised ROC curves with various values of w on the total dataset. Note that $w = 1$ corresponds to the model of VP only. We can see that the model of VP only (ROC with $w = 1$) shows high true positive rate at high saliency threshold ($> 10\%$), but shows a weak ability of fixation prediction at low saliency threshold ($< 10\%$). This means that most human fixation points distribute close to the vanishing point, which is consentient with the result of the psychophysical experiment in Section II. However, the VP-only model misses important fixation points which are far from the vanishing point, which results in low ROC below 10%

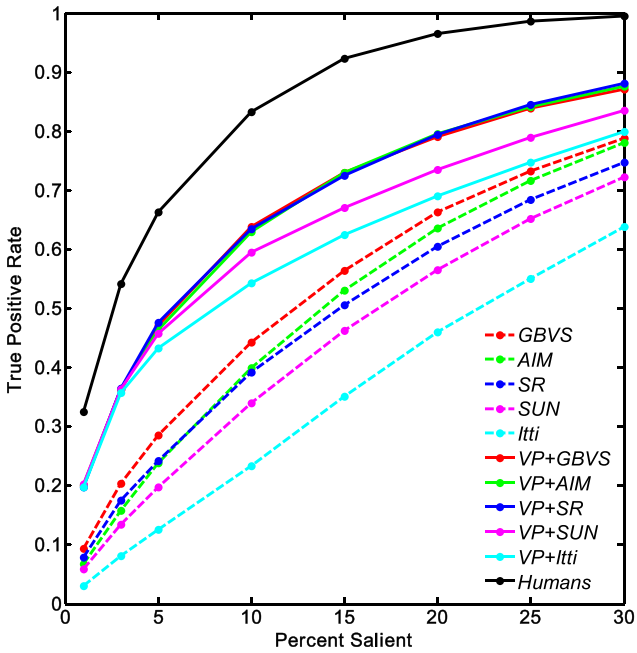


Fig. 11. The revised ROC curves of different algorithms. The solid lines show the ROCs of the algorithms combined with the VP information, and the dashed lines show the ROCs of the classical saliency algorithms.

saliency threshold. The result also reveals that both bottom-up and top-down information are necessary for fixation prediction in traffic scenes, and vanishing point is an important top-down control for guiding visual attention of drivers. Finally, we select $w = 0.8$ as the optimal weight for combining the bottom-up and top-down information in Equation (1).

Fig. 11 shows comparisons of the revised ROC curves of the five bottom-up models and their corresponding VP information combined ones. We can find that the models combined with the VP feature perform much better than the original models with bottom-up feature alone. For example, at the 10% salient location threshold, the Itti model with the VP feature performs at 0.54 and the original Itti performs at 0.22, which demonstrates a remarkable improvement in performance. Generally, the true positive rate of the saliency model with the VP information has improved about 0.1–0.35, which means that our model predicts human fixation points more quickly and precisely at low saliency threshold.

D. Normalized Scan-Path Saliency

To further quantify how well our model's prediction matches the subjects' eye positions, we also employ the Normalized Scan-path Salience (NSS) [26], [51]. The NSS scores were extracted at each of the multiple fixation point along a subject's scan-path, and the mean of these values, called the NSS, was taken as a measure of the correspondence between the saliency map and scan-path [51]. In this paper, we define the NSS according to [7], [26], as follows:

$$\mathrm{NSS} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma_S}\left(S(x_i, y_i) - \mu_S\right)\right) \tag{2}$$

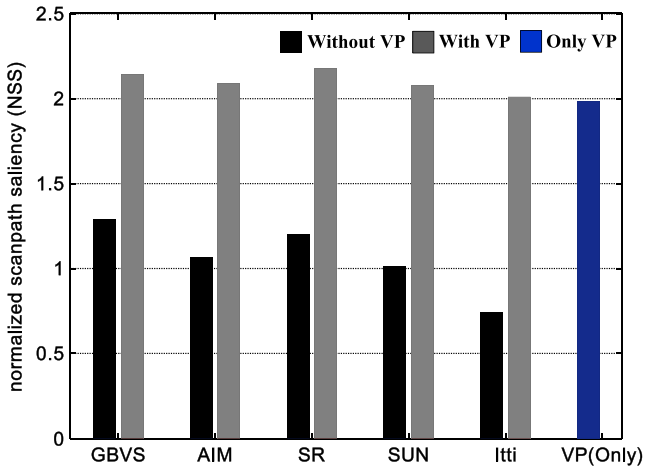


Fig. 12. Normalized scan-path saliency (NSS) scores from comparing different models' predictions with actual eye positions.

where μ_S and σ_S are respectively the mean and standard deviation of S, and S is the model's saliency map, (x_i, y_i) is the position of fixation point, N is the number of fixation points. NSS $= 1$ indicates that the subject's eye position falls within a region where predicted density is one standard deviation above average while NSS $= 0$ means that the model performs at chance level [7].

In Fig. 12, the black bars show the NSS scores of the bottom-up saliency models without VP information, the blue bar shows the score of VP only, and the grey bars show the scores of the models combined VP. We can clearly see that there is a significant difference between the models with and without VP. This figure indicates that the algorithmic performance of the bottom-up saliency models (GBVS [18], AIM [20], SR [6], SUN [21], Itti [13]) are remarkably improved by combining with VP information as a kind of top-down control. The results also mean that the models combined with VP information can predict human fixations more robustly. Similar with the conclusion from the revised ROC evaluation, this figure also shows that the model with VP only obtains the lower NSS score than almost all VP-based top-down models, which indicates that it is not enough to predict fixations with only VP information. On the other hand, the NSS score of VP only is higher than all considered classical bottom-up saliency models, which means that the VP feature is an important top-down source for visual attention in traffic driving environment.

V. Discussion and Conclusion

Previous studies have proposed that two main categories of attention (endogenous and exogenous attention) are involved in driving [42], [52], [53]. Exogenous attention (bottom-up) would be attracted by specific objects such as traffic signs and lights or prompted by sudden changes in the visual field, such as another road moving into the field of view. Endogenous attention (top-down) would be controlled by the driver's knowledge of the current road and traffic conditions, including the awareness of sources of important information, such as hazards, and could anticipate problems that are about to arise. Konstantopoulos *et al.*

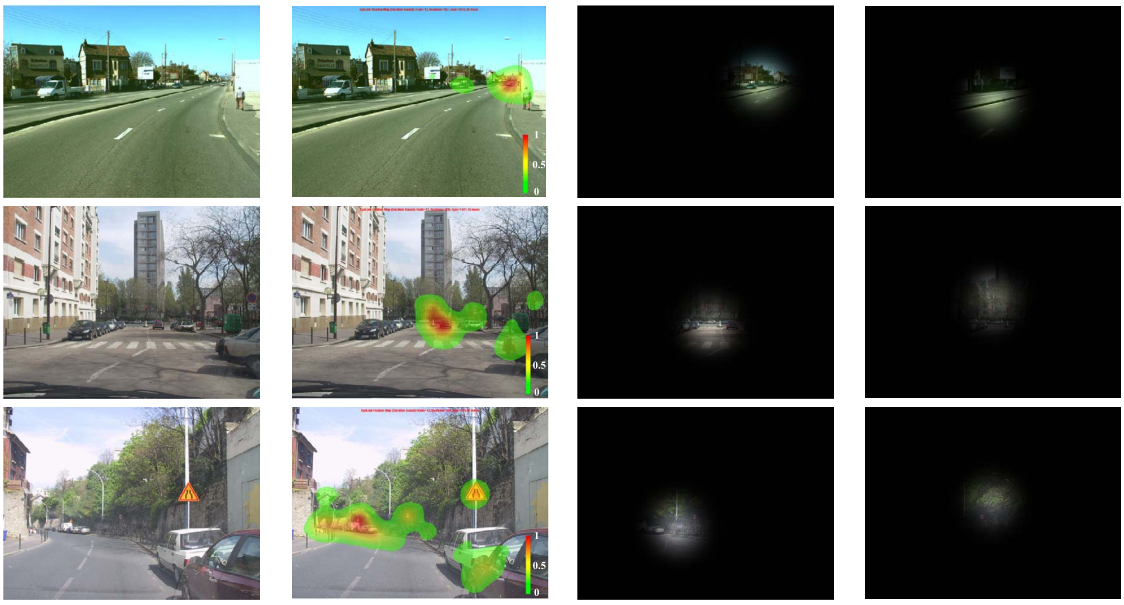


Fig. 13. Comparison of the center bias model and our proposed model in the curved road environment. (a) Original image. (b) human saliency map. (c) VP based saliency map. (d) saliency map with center bias.

concluded that driver's visual attention is a function of driving experience and visibility [36]. Experienced drivers can allocate the two categories of attention optimally, whereas new drivers might fail to do so. On the one hand, experienced drivers clearly know that the current driving road is the main information source from which the drivers analyze the traffic conditions and anticipate potential problems. Therefore, in a normal traffic driving environment, experienced drivers' endogenous attention, namely top-down control, can be mainly allocated to the end of the road in front of the vehicle. This hypothesis is supported by the behavioral results that the driver group fixates most of their gazes on the vanishing point of the road. A similar result was reported by Underwood *et al.*, who found that the road far-ahead and the road mid-ahead attract more fixations than any other part of the scene and that the road near-left and the road near-right tend to attract fewer fixations than the other parts of the scene [42]. On the other hand, by fixating the vanishing point of the road, drivers can also have a larger and better visual field to cope with emerging accidents [54]. We deduce that regarding the vanishing point of the road to be a top-down control or endogenous attention is consistent with the visual attentional mechanism in a traffic environment. Our results also show that the VP-based top-down models have better performances in traffic saliency detection than the classical bottom-up models.

The contributions of this paper can be summarized as follows. 1) The eye-movement data were collected and analyzed by our psychophysical approach to the traffic driving environment. The bottom-up and top-down attentional mechanisms were considered in the behavioral experiment. 2) A top-down attention guided framework was designed to detect the traffic saliency maps. The previous models of Judd [17] and Borji *et al.* [7], [55] work well in some specific scenes or tasks but are difficult to work successfully in the traffic environment because traffic scenes are quite different from normal scenes. 3) The vanishing point of the road was proposed for top-down control or endogenous attention in traffic saliency detection. As far as we know, although traffic-related behavioral experiments have shown that the road far ahead attracted more of a driver's fixations than any other part of the scene, this important information has not been applied to predict the driver's eye movements and traffic saliency detection. Our results show that the model, which combines bottom-up saliency detection with top-down attentional control, can effectively estimate a human's actual eye gaze and salient processing in the traffic driving environment, indicating a potential application in guiding the initiative traffic saliency gazes in unmanned vehicle systems. 4) A more accurate ground-truth was anew defined in our work. We considered the different weights of the fixations in the saliency area based on the fixation durations. So, the new ground-truth contains more intensive information, rather than binary fixation points. The revised ROC curve is more convincing and more accurate when benchmarking on the revised ground-truth weighted by the fixation durations.

People may argue that our model is similar to the center bias methods that were proposed by several researchers [17], [25], [56] in recent years. Here, we note that our model is quite different from their principle. The center bias methods could be effective if the road is in the center of the visual field. However, they cannot work when the road is curved or the front of the road is out of the center of the scene. Fig. 13 illustrates the comparison of the center bias model and our model in the curved road environment. This figure shows that the center bias model is not suitable for the curved road and other complex situations, such as when the road is not in the center of the scene.

In conclusion, in this paper, based on previous studies of eye movement in driving and our behavioral experimental results, we found that the drivers' attention mostly focuses on the end of the road in front of the vehicle, where it overlap with

the vanishing point of the road, on most occasions. Then, we proposed that the vanishing point of the road can be regarded as a valuable kind of top-down guidance in the traffic saliency detection model. Subsequently, we proposed a framework of a classical bottom-up and VP-based top-down combined traffic saliency detection model. Finally, the classical ROC, AUC, the revised ROC and NSS measures were applied to evaluate the performance of the proposed strategy, and the results showed that our method can effectively simulate the attentive areas in a traffic environment compared with those classical bottom-up saliency models.

However, there also are some limitations in our current work. In the traffic driving environment, many unpredictable and uncontrollable factors such as hazy sky, barrier, pedestrians, and traffic accidents and so on may affect the driver's attention and visual saliency. For example, in a crowed road, especially when there are some obstacles in the front of the vehicle, the driver's eye movements may be different and he/she may admit different behaviors. However, although our model cannot solve all kinds of situations in driving, we give a framework of a classical bottom-up and top-down combined model. Suppose the situation when there are some obstacles in the front of the vehicle, the top-down VP information may be interfered, but the low-level attributes of input images would still play a role in the model. The saliency can also be detected to a large extent. In the future, the most important work is to improve our model by analyzing dynamic video to fit the real driving situations. Furthermore, more scene analyzing technologies such as scene segmentation, object detection, machine learning and so on are needed to develop to account for other important traffic situations. As a specific issue, the algorithm for the vanishing point detection should be improved to detect the vanishing point in a curved road both robustly and effectively, and the computation time should be reduced.

ACKNOWLEDGMENT

The authors would like to thank Prof. C.-Y. Li for the valuable suggestions and H. Kong and his colleagues for their source codes for the vanishing-point detection.

REFERENCES

[1] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vis. Res.*, vol. 45, no. 2, pp. 205–231, 2005.

[2] S. Frintrop, *Vocus: A Visual Attention System for Object Detection and Goal-Directed Search*. New York, NY, USA: Spriger-Verlag, 2006.

[3] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, 2006.

[4] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 300–312, Feb. 2007.

[5] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *IEEE Trans. Robot.*, vol. 25, no. 4, pp. 861–873, Aug. 2009.

[6] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2007, pp. 1–8.

[7] A. Borji, D. N. Sihite, and L. Itti, "What/where to look next? Modeling top-down visual attention in complex interactive environments," *IEEE Trans. Syst., Man, Cyb., Syst.*, vol. 44, no. 5, pp. 523–538, May 2014.

[8] M. Hayhoe and D. Ballard, "Modeling task control of eye movements," *Curr. Biol.*, vol. 24, no. 13, pp. R622–R628, 2014.

[9] D. E. Broadbent, "The role of auditory localization in attention and memory span," *J. Exp. Psych.*, vol. 47, no. 3, p. 191, 1954.

[10] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psych.*, vol. 12, no. 1, pp. 97–136, 1980.

[11] A. L. Yarbus, B. Haigh, and L. A. Rigss, *Eye Movements and Vision*, vol. 2. New York, NY, USA: Plenum, 1967, no. 5.10.

[12] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," in *Matters Intelligent*. New York, NY, USA: Springer-Verlag, 1987, pp. 115–141.

[13] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[14] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vis. Res.*, vol. 40, no. 10, pp. 1489–1506, 2000.

[15] L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.

[16] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.

[17] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2009, pp. 2106–2113.

[18] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Neural Inf. Process. Syst.*, 2006, pp. 545–552.

[19] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *Proc. Neural Inf. Proc. Syst.*, 2009, pp. 681–688.

[20] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Proc. Neural Inf. Proc. Syst.*, 2005, pp. 155–162.

[21] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "Sun: A Bayesian framework for saliency using natural statistics," *J. Vis.*, vol. 8, no. 7, p. 32, 2008.

[22] M. F. Land and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?" *Vis. Res.*, vol. 41, no. 25, pp. 3559–3565, 2001.

[23] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends Cogn. Sci.*, vol. 7, no. 11, pp. 498–504, 2003.

[24] R. D. Rimey and C. M. Brown, "Control of selective perception using Bayes nets and decision theory," *Int. J. Comp. Vis.*, vol. 12, no. 2/3, pp. 173–207, 1994.

[25] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *J. Vis.*, vol. 11, no. 3, p. 9, 2011.

[26] R. J. Peters and L. Itti, "Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2007, pp. 1–8.

[27] R. Peters and L. Itti, "Congruence between model and human attention reveals unique signatures of critical visual events," in *Proc. Neural Inf. Proc. Syst.*, 2007, pp. 1145–1152.

[28] J. H. Lim, Y. Liu, and O. Tsimhoni, "Investigation of driver performance with night-vision and pedestrian-detection systems—Part 2: Queuing network human performance modeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 765–772, Sep. 2010.

[29] A. De La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE Trans. Ind. Electron.*, vol. 44, no. 6, pp. 848–859, Dec. 1997.

[30] S. Gupte, O. Masoud, R. F. Martin, and N. P. Papanikolopoulos, "Detection and classification of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 1, pp. 37–47, Mar. 2002.

[31] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1484–1497, Dec. 2012.

[32] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vis. Res.*, vol. 43, no. 9, pp. 1035–1045, 2003.

[33] D. Miniotas and B. Velichkovsky, "Eye movements and attention during simulated driving," *Elektr ir Elektrotechnika*, vol. 43, no. 1, pp. 7–12, 2003.

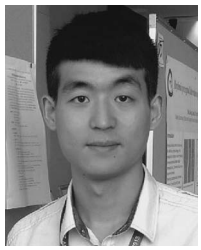
[34] M. Ko, L. Higgins, S. T. Chrysler, and D. Lord, "Effect of driving environment on drivers' eye movements: Re-analyzing previously collected eye-tracker data," in *Proc. Transp. Res. Board 89th Annu. Meet.*, 2010, pp. 1–15.

[35] P. Green, *Where do Drivers Look While Driving (and for how Long)*. Tucson, AZ, USA: Hum. Factors Traffic Safety, 2002, pp. 77–110.

[36] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving," *Accident Anal. Prev.*, vol. 42, no. 3, pp. 827–834, 2010.

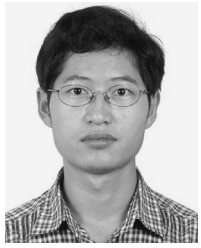
[37] T. Deng, E.-Q. Luo, Y.-S. Zhang, and H.-M. Yan, "Selective attention-based saliency of traffic images and characteristics of eye movement," *J. Univ. Electron. Sci. Techn. Chin.*, vol. 43, no. 4, pp. 624–628, 2014.

[38] H. Kong, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2009, pp. 96–103.
[39] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," in *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2211–2220, Aug. 2010.
[40] T. Deng, A. Chen, M. Gao, and H. Yan, "Top-down based saliency model in traffic driving environment," in *Proc. IEEE ITSC*, 2014, pp. 75–80.
[41] F. Baluch and L. Itti, "Mechanisms of top-down attention," *Trends Neur.*, vol. 34, no. 4, pp. 210–224, 2011.
[42] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall, "Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers," *Ergonomics*, vol. 46, no. 6, pp. 629–646, 2003.
[43] P. Moghadam, J. A. Starzyk, and W. S. Wijesoma, "Fast vanishing-point detection in unstructured environments," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 425–430, Jan. 2012.
[44] M. Nieto and L. Salgado, "Real-time vanishing point estimation in road sequences using adaptive steerable filter banks," in *Advanced Concepts for Intelligent Vision Systems*. New York, NY, USA: Springer-Verlag, 2007, pp. 840–848.
[45] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2004, vol. 1, pp. 470–477.
[46] C. Rasmussen, "Texture-based vanishing point voting for road shape estimation," in *Proc. BMVC*, 2004, pp. 1–10.
[47] H. Kong, S. E. Sarma, and F. Tang, "Generalizing Laplacian of Gaussian filters for vanishing-point detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 408–418, Mar. 2013.
[48] D. M. Green *et al.*, *Signal Detection Theory and Psychophysics*. New York, NY, USA: Wiley, 1966, vol. 1.
[49] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
[50] J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao, "Predicting human gaze beyond pixels," *J. Vis.*, vol. 14, no. 1, p. 28, 2014.
[51] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vis. Res.*, vol. 45, no. 18, pp. 2397–2416, 2005.
[52] G. Underwood, D. Crundall, and P. Chapman, "Selective searching while driving: The role of experience in hazard detection and general surveillance," *Ergonomics*, vol. 45, no. 1, pp. 1–12, 2002.
[53] G. Underwood, P. Chapman, K. Bowden, and D. Crundall, "Visual search while driving: Skill and awareness during inspection of the scene," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 5, no. 2, pp. 87–97, 2002.
[54] J.-G. Yao, X. Gao, H.-M. Yan, and C.-Y. Li, "Field of attention for instantaneous object recognition," *PloS One*, vol. 6, no. 1, 2011, Art. no. e16343.
[55] A. Borji, "Boosting bottom-up and top-down visual features for saliency estimation," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2012, pp. 438–445.
[56] B. W. Tatler, "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *J. Vis.*, vol. 7, no. 14, p. 4, 2007.



Tao Deng is currently working toward the Ph.D. degree in biomedical engineering from University of Electronic Science and Technology of China, Chengdu, China.

His research interests include visual attention, cognition, vision computation, and saliency detection.



Kaifu Yang received the B.Sc. and M.Sc. degrees in biomedical engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2009 and 2012, respectively, where he is currently working toward the Ph.D. degree.

His research interests include visual mechanism modeling and image processing.

Yongjie Li (M'14) received the Ph.D. degree in biomedical engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004.

He is currently a Professor with Key Laboratory for Neuroinformation of the Ministry of Education, School of Life Science and Technology, UESTC. His research interests include visual mechanism modeling, image processing, and intelligent computation.

Hongmei Yan received the M.S. and Ph.D. degrees from Chongqing University, Chongqing, China, in 2000 and 2003, respectively.

She is currently a Professor with Key Laboratory for Neuroinformation of the Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include visual cognition, eye movements, visual attention, and saliency detection.