Bayes Saliency-Based Object Proposal Generator for Nighttime Traffic Images

Hulin Kuang, Kai-Fu Yang, Long Chen, *Member, IEEE*, Yong-Jie Li, *Senior Member, IEEE*, Leanne Lai Hang Chan, *Member, IEEE*, and Hong Yan, *Fellow, IEEE*

***Abstract*— Object proposal is one of the most key pre-processing steps for nighttime vehicle detection systems in intelligent transportation systems. However, most current object proposal methods are developed on daytime data sets, and these methods demonstrate unsatisfactory results when they are used on nighttime images. Therefore, this paper presents a novel Bayes saliency-based object proposal generator for nighttime RGB traffic images to generate a modest and accurate set of proposals, which are more likely to be vehicles for preceding vehicle detection. First, we propose a new Bayes saliency detection approach in which prior estimation, feature extraction, weight estimation, and Bayes rule are used to compute saliency maps. Then, we propose a simple but effective object proposal generator based on the Bayes saliency map. Multi-scale sliding window, proposal rejecting, scoring, and non-maximum suppression are combined to generate a modest and effective set of proposals. Experimental results demonstrate that our proposed approach generates a modest set of proposals and outperforms some state-of-the-art methods on nighttime images in terms of various evaluation metrics. Furthermore, our proposed object proposal approach can improve the detection performance and the speed of several state-of-the-art vehicle detection approaches.**

***Index Terms*— Object proposal, saliency detection, Bayes rule, nighttime vehicle detection.**

I. Introduction

IN RECENT years, object proposal, a technique that generates a set of proposals (windows or regions) which are more likely to be the attended objects, has become an important and useful topic in computer vision [1]. The object proposal methods generate fewer windows than the traditional sliding window mechanism, thus they can speed up many high-level tasks in computer vision such as object recognition and detection.

About 32% of all vehicular accidents is caused by rear-end collisions that are one of the most fatal kind of traffic accident [2]. Most of these accidents occur at night. Thus nighttime preceding vehicle detection is very important for road safety in advanced driver assistance systems (ADAS) and autonomous driving systems (ADS) [3]. Extracting regions that contain most preceding vehicles is a key pre-processing step of nighttime preceding vehicle detection.

Current object proposal approaches process daytime images such as Imagenet [4] and Visual Object Classes (VOC) [5] datasets where the images have eminent lightness and the attributes of objects are very clear. However, in nighttime images, due to the low lightness and the low contrast the attributes of vehicles such as color, texture and edges become unclear [3], leading to the unsatisfactory performance using current object proposal approaches. Besides, most object proposal methods depend on methods which take additional computation time such as learning techniques [1], [6], [7]. The objective of this paper is to propose an efficient object proposal generator which is effective to locate vehicles in nighttime traffic images.

Saliency detection can be considered as a type of object proposal approaches where proposals are sets of salient pixels. However, for tasks such as vehicle detection we pay more attention to rectangular proposals which are windows that contain vehicles rather than pixels. To design an efficient algorithm that generates a modest set of proposals that can locate the salient vehicles in RGB images, we extract the coarse saliency map as the basis of generating accurate object proposals. The framework of the proposed method is shown in Fig. 1.

There are two major steps: Bayes saliency detection and object proposal generation based on the saliency map. For Bayes saliency detection, we first estimate the probability prior of a pixel belonging to the salient objects and the background via edge detection on the input nighttime image. Meantime, we extract three features: luminance, local contrast and vehicle taillight map which robustly represent the attributes of vehicles within the nighttime image. To compute the condition likelihood functions of the observed background and potential salient pixels, we utilize steps as follows. To divide all pixels into background set and salient pixel set, we binarize the prior map by searching the optimal threshold which separates the background from the possible salient pixels in each feature

Manuscript received July 16, 2016; revised March 8, 2017 and April 30, 2017; accepted May 6, 2017. Date of publication June 2, 2017; date of current version March 7, 2018. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project CityU 123412; in part by the City University of Hong Kong under Project 9610308; in part by the Major State Basic Research Program of China under Project 2013CB329401; and in part by the Natural Science Foundations of China under Project 91420105. The Associate Editor for this paper was L. M. Bergasa. *(Corresponding authors: Yong-Jie Li; Leanne Lai Hang Chan.)*

H. Kuang, L. L. H. Chan, and H. Yan are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: hlkuang2-c@my.cityu.edu.hk; leanne.chan@cityu.edu.hk; h.yan@cityu.edu.hk).

K.-F. Yang and Y.-J. Li are with the Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: yangkf@uestc.edu.cn; liyj@uestc.edu.cn).

L. Chen is with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China (e-mail: chenl46@mail.sysu.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.2017.2702665

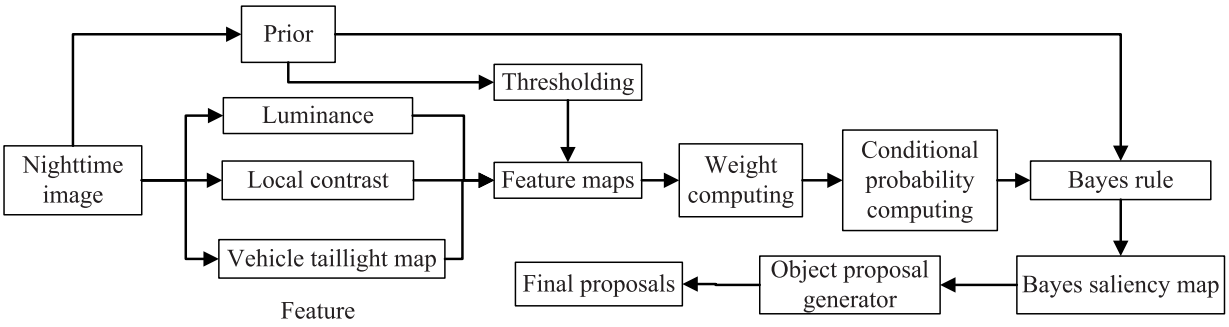


Fig. 1. The framework of the proposed object proposal approach.

map. Weights of each feature for each class (background and salient pixels) are estimated using variance of feature of each class. The condition probabilities are computed by multiple probabilities of each feature with weights. Using Bayes rule, we obtain the posterior probability of a pixel belong to the salient pixels (i.e., the saliency map). After Bayes saliency detection, we propose a new object proposal generator based on the obtained Bayes saliency map. Multi-scale sliding window mechanism, proposal rejecting (i.e., rejecting object proposals that are less likely to be objects) based on saliency map, score computing and non-maximum suppression (NMS) are utilized to obtain a modest and accurate set of proposals to locate vehicle regions in nighttime traffic images.

The contributions of this paper are summarized as follows. (1) A novel and fast Bayes saliency detection which detects salient pixels within regions of vehicles is proposed. Multiple features of vehicles are integrated with carefully designed weights using probability theory. (2) An effective object proposal approach is present to relate the saliency map with object proposals and generate proposals with high detection rate. Besides, simple proposal rejecting is used to reduce the number of proposals.

The rest of the paper is organized as follows. Related work is described in Section II. The details of the proposed object proposal approach are introduced in Section III. The experimental results are illustrated in Section IV. Finally, Section V states the conclusions.

II. Related Work

Object proposal (sometimes called "detection proposal" or "objectness") is a technique that produces a set of windows or regions which are likely to contain objects of any category, to speed up the traditional sliding window, and has been well reviewed [8], [9].

Several previous methods trained a scoring or ranking model using cues with traditional learning techniques such as Bayes framework and support vector machine (SVM) to compute the scores that denoted the likeliness of a window to be an object [1], [6], [7], [10]–[12]. Alexe *et al.* [1] combined multiple cues in a Bayes framework and learnt a model to compute objectness of each window. Rahtu *et al.* [6] proposed a SVM-cascade framework to compute and calibrate scores of each window. This SVM-cascade framework was followed by BING [10], a very fast approach, where a very simple gradient feature was utilized. Endres and Hoeim [7] trained a ranking model to re-order the scores. In [11], the object proposals were the solution of a binary segmentation problem, constrained parametric min-cut problems (CPMC) with various initial seeds and a ranking model was learnt using several mid-level region cues. Lee *et al.* [12] introduced a novel structured learning approach, parametric min-loss (PML), to learn how to combine several mid-level cues for generating a small set of object proposals with high detection rate. The above mentioned methods require one to retrain models when new datasets are used, which lengthens the processing time.

Some methods can be directly used on customized datasets such as [13] and [14]. In SelectiveSearch [13], features and superpixel similar functions were carefully designed to deliberately merge low-level superpixels. EdgeBoxes [14] firstly extracted accurate object edge estimation of each image. The objectness score of each box, which was the likelihood of a box to be an object, was obtained by subtracting edges that overlapped the boxâŁ™s boundaries from the number of contours that were contained by the box. SelectiveSearch obtains high recall on daytime images but is very slow. EdgeBoxes is one of the best object proposal methods but it might omit some vehicles in nighttime images if only a small part of proposals are used.

Most of the current object proposals are based on segments. Manen *et al.* [15] used similar features in [13] but the weights of the merging function were learnt to combine low-level superpixels under a randomised merging process. Rantalankila *et al.* [16] improved SelectiveSearch [13] by using different features and similar functions when merging superpixels and utilized a CPMC-like process to obtain the final proposals. An approach, called multiscale combinatorial grouping (MCG), combined multi-scale hierarchical segmentation with object proposal generation in a unified framework and developed an improved cuts algorithm and a different merging strategy [17]. Various low-level superpixels and features were utilized by RIGOR [18] to merge segments. Besides, two speedup techniques were designed to avoid redundant computations. Xiao *et al.* [19] combined a novel distance metric in highly complex scenarios with some existing distance metrics for low-complexity cases to merge segments with different complexities. Wang *et al.* [20] proposed a novel multi-branch hierarchical segmentation approach which learnt multiple grouping strategies in each step to correct errors in

early stages. These low-level segments based methods might not be suitable for nighttime images because the segments in nighttime images may not be as accurate as segments within daytime images.

Recently several techniques which are used to calibrate some existing object proposals methods are proposed [21]–[24]. Chen *et al.* [21] proposed the multi-thresholding straddling expansion (MTSE) approach to reduce localization bias and avoid time-consuming diversification strategies. This method was a proposal calibration process and could be combined into any existing methods. Inspired by the fact that a good object proposal should bound the optimized closed contour. Lu *et al.* [22] defined a new closed contour measure using closed path integral to dismiss a lot of false proposals whose closed contours were not tightly bounded. Liu *et al.* [23] proposed an uncomplicated box aggregation function via analyzing the statistical properties of proposals which contained objects for proposal decimation. In [24], Wang and Shen reported a new ranking model which divided proposals into two subsets and considered partial ranking constraints to generate a set of the top-k proposals with high detection rate. Kuang *et al.* [3] combined vehicle light detection with EdgeBoxes [14] to extract accurate regions-of-interest (ROIs) for nighttime vehicle detection. The performance of these methods depends on the existing object proposal approaches.

With the development of deep learning technologies, researchers have applied deep learning into object proposals [25]–[27]. Some objects may be omitted using object proposal approaches. To solve this problem, a boosting approach which made full use of the hierarchical convolutional neural network (CNN) features was proposed in [25] for identifying regions containing objects rapidly. DeepBox in [26] utilized a new four-layer CNNs to efficiently rerank proposals under a bottom-up framework. In [27] Ghodrati *et al.* designed a CNN-based method and built an inverse cascade to choose the most potential object locations and calibrate these proposals in a coarse-to-fine manner. Faster R-CNN designed a region proposal network (RPN) to generate a set of object proposals [28]. CNN-based object proposals are effective but the speed of the CNN is limited without using the graphics processing unit (GPU).

There are some works that propose different techniques based on headlight or taillight detection for obtaining a set of potential vehicle candidates for different applications (such as vehicle detection and tracking automatic headlamp controller) in nighttime driving conditions [29]–[39]. An overview of on-road night-time vehicle light detection using camera sensors was given in [34]. The common pipeline of vehicle light detection methods includes image processing, light region detection, pairing or classification and tracking. We divide these methods into two types: learning-based methods and thresholding-based methods.

Typical learning-based methods are summarized as follows. To detect the headlights of a vehicle, a method based on Intelligent Headlight Control (IHC) systems was proposed in [29], which combines the spot finder, multiple classifiers for different types of lights and OR gate for fusing the different classification results. Nighttime vehicle detection based on headlights or taillights were proposed in [30] for the automatic headlight control. A classifier based on different appearance features extracted from a novel image sensor was trained to detect vehicle lights, and then the resemblance to a vehicle light was conducted using a novel temporal coherence analysis. In [33], an effective system was presented for detecting vehicles in order to automatically change the vehicle headlights between low and high beams. Some candidate bright objects were selected via bright object extraction based on adaptive thresholding, clustering and tracking based on Kalman filtering. Then a classifier was trained with support vector machine to separate vehicle lights from road signs reflections or nuisance artifacts. In [36], a learning-based light detection at night was proposed. Firstly, thresholding and connected component analysis were applied to find bright spots. Then features were extracted from these bright spots for training a classifier of support vector machine or Adaboost, which was used to classify these bright spots into headlights, taillights, streetlights and others. The method named VeDANt (Vehicle Detection using Active-learning during Nighttime) proposed in [39] used an improved active learning to train Adaboost classifiers with Haar-like features of gray-level input images to obtain taillight candidate windows, which were then further verified using perspective geometries and color information of the taillights.

There are also some thresholding-based methods for vehicle light detection. For example, O'Malley *et al.* [31] detected and tracked vehicle rear-lamp pairs by first using a camera-configuration process to optimize the appearance of rear lamps, followed by the segmenting of rear lamps with a red-color threshold and then the paring of lamps using color cross-correlation symmetry analysis. To detect vehicles at night, Chen [32] proposed a method to locate headlights and rear-lights by first separating the bright objects (i.e., headlights and taillights) from the nighttime images by using fast automatic multilevel thresholding and then detecting the vehicles by processing these bright objects using a knowledge-based connected-component analysis procedure. Wang *et al.* [35] used the brightness of the taillights as features and then an improved Otsu method was utilized to detect the taillights based on the cumulative histogram and adaptive thresholds for the area and redness of bright spots' areas, finally the taillight candidates were paired. Eum and Jung [37] combined low-exposure (LE) image blob detection with auto exposure (AE) image blob detection. Fixed thresholds were used. For the near taillights and the near and distant headlights, a fixed threshold obtained from LE images was used. For the distant taillights, a LoG filter threshold obtained from AE images was used. In [38], vehicle candidates were detected by combing lane detection where lane markers, brightness, slenderness and proximity were used to detect the positions of lane markers and vehicles by integrating the taillight standing-out process, adaptive thresholding, centroid detection and taillight pairing.

In most of the methods mentioned above, each vehicle candidate is shown by a bounding box that contains the detected lights. From the results reported, these methods have been proven to be useful for generate a set of vehicle

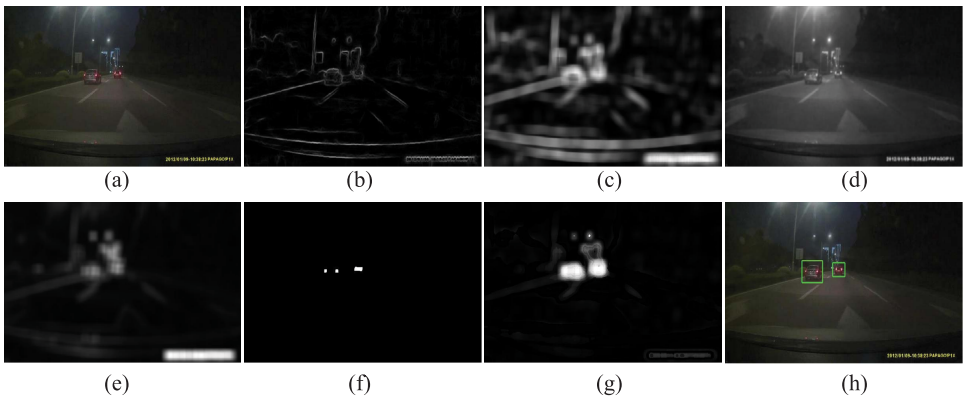


Fig. 2. Examples of all steps of the proposed method. (a) The input nighttime traffic images. (b) The edge map obtained by [40]. (c) demonstrates the $p(s)$ estimated by our prior estimation. (d)-(f) denotes the examples of luminance feature, local contrast and vehicle taillight map respectively. (g) represents the Bayes saliency map. (h) shows the proposals obtained by our proposed approach.

candidates. However, they are all based on vehicle light detection, the performance of which is influenced by road scene, camera parameters, reflection and so on. Therefore, some vehicle candidates may be missed. To detect all preceding vehicles in front of the driver, we proposed a saliency detection based object proposal method, which also applies vehicle light detection, but utilizes new features, Bayes rule and object generating based on multiscale sliding window to correct errors from vehicle light detection.

III. The Proposed Object Proposal Generator

As introduced in Section 1, our proposed object proposal approach has two major steps: Bayes saliency detection and saliency based object proposal generation. The details are stated in the following sections.

A. Bayes Saliency Detection

1) Prior Estimation: From observations, we find that though the nighttime traffic images have low lightness and low contrast, edges are noticeable in salient objects (see Fig. 2(b)). The surrounding pixels of edges are more likely to belong to salient objects. To estimate priors, we first compute the edge map of the input nighttime image using the method in [40] which can extract accurate edge maps. Considering the surrounding pixels of edges, the edge map is calibrated by convoluting with an average filter whose size is $S_f = 21 \times 21$ as

$$P(x, y) = E(x, y) \otimes A(x, y), \tag{1}$$

where $E(x, y)$ is the edge value of pixel (x, y), $A(x, y)$ is an average filter, $\otimes$ denotes the convolution operator and $P(x, y)$ is the calibrated edge map. In this paper, we design Gaussian filters as follows: for a fixed mask size such as $S_f = 21 \times 21$, we design the σ (i.e., the standard deviation of Gaussian function) such that the values of the Gaussian function that are outside the mask is close to zero. Thus, we can use the mask size to denote the size of the Gaussian filter.

The calibrated edge map $P(x, y)$ is normalized to the range [0, 1] and vectorized to be the final estimated prior probability of a pixel belonging to the salient pixels ($p(s)$). According to the probability theory, the probability of a pixel being background pixel ($p(b)$) is equal to $1 - p(s)$. Fig. 2(c) demonstrates an example of $p(s)$. We find that the prior estimation is effective because the pixels in the two vehicles are white (close to 1) while most of background are black. Note that the prior map in our proposed method does not have to be very accurate. It is enough for the prior map to include the most edges of vehicles and provide information on the rough locations of the vehicles, since the inaccuracy of the prior map, for example, of a relatively large vehicle, will be corrected as much as possible by our other strategies described below, for example, the three types of features, the Bayes rule and multi-scale sliding window based object proposal generation.

2) Feature Extraction: The salient objects in nighttime often have bright regions. And the contrast of object regions are often higher than the contrast of background regions. Besides, at night vehicles often turn on the taillight which are very salient and useful for distinguishing vehicles from the background. Thus, to represent the differences between the salient vehicles and the background pixels, we extract three types of features: luminance, local contrast and vehicle taillight map.

The luminance image is computed as the average of the R, G, B channels of the input image, that is, $f_{lum} = (R + G + B)/3$ (see an example in Fig. 2(d) where several regions (e.g., lights) within the vehicle region are bright). The local contrast f_{con} of each pixel is computed as the normalized variance of gray values within a $S_w = 7 \times 7$ window centered at this pixel. An example is given in Fig. 2(e) where we find that the local contrast of the two vehicles is brighter (higher) than the background. The luminance and local contrast features are smoothed using an Gaussian average filter whose size is $S_f = 21 \times 21$ and they are also normalized to [0,1].

The computation process of the vehicle taillight map is similar to the process in [3] and described below. First, we convert the RGB nighttime images to intensity images and reduce noise using an empirical threshold (0.4 referred from [41]). Second, the Nakagami images [3], [41] are

estimated by utilizing a sliding window mechanism (window size: $S_w = 7 \times 7$ pixels). Note that the Nakagami image presents the local Nakagami distribution at each pixel, and it was found in [41] that the Nakagami distribution can well model the characteristics (e.g., the scattering property) of vehicle light regions. This makes the local Nakagami parameter at each pixel an appropriate index to differentiate the vehicle light regions from others.

Subsequently, we detect possible vehicle taillights by the optimal thresholds. Because our proposed method is a key preprocessing step of some high-level tasks such as preceding vehicle detection, we should make sure all interested vehicles are contained by the object proposals, which means we should detect all taillights. After computing the Nakagami images of all positive samples in the dataset, we try low-level and high-level thresholds close to the maximum and minimum values respectively to segment the Nakagami images. The values between the low- and high-level thresholds are set to 0 and other values are set to 1. If a pair of thresholds can detect all taillights (most of pixels within taillights are retained) in all positive samples and detect minimum non-taillight regions, we choose them as the optimal thresholds (the low-level threshold is $L_{th} = 0.08$ while the high-level threshold is $H_{th} = 11.6$ in this work).

Then, we compute the red pixel rate of each pixel: the rate of the number of red pixels in a $S_w = 7 \times 7$ window centered at this pixel and the number of all pixels in the window, and the thresholding result is multiplied by the red pixel rate to obtain the coarse vehicle taillight map. Because the pixels which are close to the detected taillight regions are more likely to be salient pixels than the pixels which are far away from the detected taillight regions, the origin vehicle taillight map is convoluted with a Gaussian filter ($S_f = 21 \times 21$) to compute the final vehicle taillight map f_{light} which is normalized to [0,1]. In Fig. 2(f), the estimated vehicle taillight map can roughly detect the taillight region within the two vehicles, thus pixels of vehicles are well represented.

3) Thresholding: To obtain conditional likelihood functions of observed salient pixels and the background, we first distinguish the possible salient regions from the background. Simply, we binarize the map of prior probability ($p(s)$) with an optimal threshold to capture rough potential salient regions and the background.

The optimal threshold T_{opt} is found by searching a possible T_k which maximizes the difference of all features between structure and background pixel sets as

$$T_{opt} = arg \max_{T_k} \sqrt{\sum_i (\frac{1}{3}(\bar{S}^i_{T_k} - \bar{B}^i_{T_k}))^2}, \tag{2}$$

where S_{T_k} and B_{T_k} denote the sets of the possible salient pixels and background respectively generated by thresholding the $p(s)$ with a certain threshold T_k, $T_k \in \{0.1, 0.12, 0.14, \ldots, 0.6\}$. The reason why we dismiss too low or too high thresholds is that the average feature value of possible salient pixels are often between 0.1 and 0.6, $i \in \{f_{lum}, f_{con}, f_{light}\}$ represents one of the three features, and $\bar{S}^i_{T_k}$ and $\bar{B}^i_{T_k}$ denote the average feature values of the salient pixels and background pixels in the i feature channel. During thresholding the weights of each feature are equal and in the following step the weights of each feature are updated.

4) Weight Computing: Different features demonstrate various abilities to distinguish salient pixels from background pixels. Therefore, we should consider the importance of each feature of each class when computing the likelihood of the observed salient pixels and background pixels. Robust features of the salient pixels should contain stable feature values (i.e., small variation). The smaller the variation of the feature belonging to a certain class, the larger the weight of the feature for this class. Then we compute the weights of each feature for the salient or background pixels as

$$w^k_i = \frac{1}{\mu} \sqrt{\frac{1 - e^{1 - var^k_i}}{1 - e}}, \quad k \in \{-1, 1\}, \tag{3}$$

where -1 and 1 denote background pixels and salient pixels respectively, $\mu = \sum_i w^k_i, i \in \{f_{lum}, f_{con}, f_{light}\}$, var^k_i is the normalized variance ($\in [0, 1]$) of feature values of the k class pixels obtained after thresholding in the i feature channel. The weights are self-adapting.

5) Conditional Likelihood Computation: After estimating the possible salient pixels and background pixels of each feature map and computing the weight of each feature for each class, we can compute the conditional likelihood of observed salient and background pixel sets. Following the kernel idea of famous Naive Bayes Classifier in machine learning [42], with the conditional independence assumption between any two features, the joint probability distribution of various features can be expressed as the product of all individual probability distributions. Thus, the observation likelihood at pixel x can be computed as

$$p(x|s) = \prod_i p(x_i|S_{T_{opt}})^{w^1_i}, \tag{4}$$

$$p(x|b) = \prod_i p(x_i|B_{T_{opt}})^{w^{-1}_i}, \tag{5}$$

where $p(x_i|S_{T_{opt}})$ and $p(x_i|B_{T_{opt}})$ are respectively the distribution functions of each feature ($i \in \{f_{lum}, f_{con}, f_{light}\}$) in the salient object and background regions. In detail, we compute observation likelihood of each feature as the normalized histogram distribution of pixels in salient or background regions in each features map. For example, $p(x_i|S_{T_{opt}})$ is computed as $N(x_i)/N(S_{T_{opt}})$, where $N(x_i)$ is the number of points in various disjoined bins containing discrete values of feature i in the region of $S_{T_{opt}}$, and $N(S_{T_{opt}})$ is the total pixel number in the region of $S_{T_{opt}}$. $p(x_i|B_{T_{opt}})$ is computed similarly. In addition, w^{-1}_i and w^1_i, which are computed using (3), indicate the contribution of each feature $i \in \{f_{lum}, f_{con}, f_{light}\}$.

6) Saliency Detection Based on Bayes Rule: With Bayes rule, the possibility of a pixel belonging to salient pixels (i.e., posterior probability and saliency map) can be computed as

$$p(s|x) = \frac{p(s)p(x|s)}{p(s)p(x|s) + p(b)p(x|b)}, \tag{6}$$

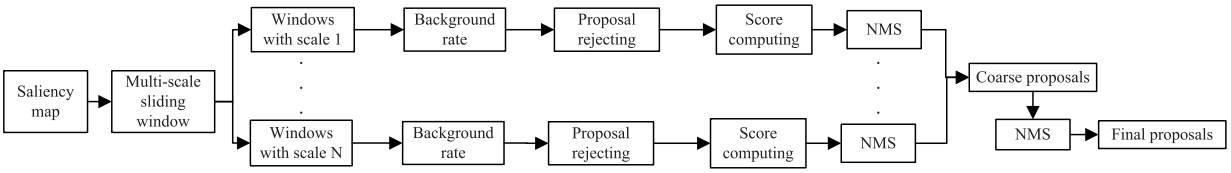


Fig. 3. The details of our proposed object proposal generator based on saliency map.

where $p(s)$ and $p(b)$ are computed during prior estimation step and the prior probabilities of a pixel belonging to salient pixels and the background, $p(x|s)$ and $p(x|b)$ are conditional likelihood functions based on the observed salient pixels and the background obtained using (4) and (5) respectively.

Bayes rule has been applied to the saliency detection before such as [43] where multiple simple features are combined using Bayes rule for saliency detection on daytime images. Different from [43] we carefully construct effective features of nighttime vehicles and design different weights of each feature for each class (the salient pixel or the background).

Afterwards, we obtain the Bayes saliency map of each nighttime image. An example is shown in Fig. 2(g) where the two vehicles are well detected by the saliency map though the boundaries of the two vehicles are missing. However, our objective is to generate a set of proposals (windows containing objects). Thus the boundaries of object are less important.

B. Object Proposal Generation

With Bayes saliency detection, we have obtained accurate saliency maps which contain most vehicles in nighttime images. How can we generate a set of object proposals simply, effectively and efficiently? Because vehicles within images can be of various sizes ranged from 20×20 to 200×200 pixels and at diversified locations, we should utilize multi-scale sliding window mechanism. From our observations, the saliency values of the vehicle regions have high average values. Besides, good proposals (windows) should contain adequate content of background around the vehicle regions, which suggests that the background rate can be used for proposal rejecting. We adopt a simple object proposal generator based on the Bayes saliency map as follows. The framework of object proposal generator is shown in Fig. 3.

First, we utilize multi-scale sliding window mechanism. The scale (width or height) of squared windows is $s \in \{20, 30, 40, \ldots, 200\}$ pixels. That is to say, there are 19 scale levels of window. Second, the background pixels rate (R_b), i.e., the rate between the number of background pixels and the number of all pixels in each window, is computed to remove windows which are less likely to contain objects. The pixels whose saliency values are lower than a threshold $T_h = 0.5$ are considered as background pixels. A good proposals should contain adequate amount of background. Therefore, only windows with $0.05 \leq R_b \leq 0.5$ are retained and other windows (proposals) are rejected (proposal rejecting). If R_b is very high, the window might be much larger than the object within it. And if R_b is too small, the window is more likely to be a part of an object.

A good window that contains the entire object and some backgrounds at the boundaries should have high score. We observe the saliency map within a good window. We find that the average saliency value in the center part of a good window subtracting the average of the rest part is higher than a bad window. Thus we compute the score of the j^{th} window using a "difference of Gaussian" (DOG) model as follows:

$$S_j = Sal_j(x, y) \otimes (G_1(x, y) - G_2(x, y)), \tag{7}$$

where S_j is the score of the j^{th} window, $Sal_j(x, y)$ is the saliency map in the j^{th} window, $G_1(x, y)$ and $G_2(x, y)$ are two averaging Gaussian filters with small size s_1 and large size s_2 respectively.

Finally, for each scale, we use non-maximum suppression (NMS) [44] which only retain the window with the highest score when a set of windows has higher overlap with each other to reduce the processing time and the number of proposals. After repeating the above steps for each scale, we generate a set of coarse proposals. By observation, we find that now one object might be surrounded by windows with different scales. Therefore, we perform NMS on the coarse proposals for generating the final set of proposals, which can avoid a situation that an object is surrounded by too many windows that have higher overlap with each other.

An example is shown in Fig. 2(h) where the green windows are proposals which have high overlaps with the groundtruths of vehicles. The generated proposals detect the two vehicles successfully. Due to proposal rejecting and NMS, our proposed object proposals approach generates a small set of proposals (at most 15 object proposals). The number of proposals of different images is different and decided by our proposed method itself (self-adaptation).

IV. Experimental Results

All experiments demonstrated in this section are conducted on a desktop computer whose processer is Intel Core i7-4770 CPU @ 3.4GHz and RAM is 8 GB. We use two datasets which are available on www.carlib.net/?page_id=35. One is the SYSU Dataset: nighttime vehicle dataset (with groundtruth). This training set includes 450 images (64×64) that contain rear of vehicle as positive samples and 2000 images without vehicles as negative samples. The test set includes 750 negative samples and 1000 positive samples. The images are taken at different sites like highways, housing estates and bridges. We use the positive samples in the test and training sets to decide the optimal threshold for vehicle taillight detection. Another one is Sun Yat-sen University Night-time Vehicle Dataset. This dataset contains 5576 images

including over 12000 nighttime vehicles in two different traffic situations, i.e., urban road with road lamp and urban road without road lamp. All these images are captured by driving recorders set on the front windshield of vehicles. We use this dataset to evaluate the proposed object proposal approach.

A. Evaluation Metrics

Referred from [9], the performance of our proposed object proposal approach is evaluated using some commonly used metrics. The first one is mean average best overlap (MABO). For each image we first compute the best (highest) overlap between the groundtruth of a certain object and all proposals, and then calculate the average of these highest overlaps (ABO). The mean of ABOs over all classes is called MABO. The higher the MABO score, the better the object proposal approach. The overlap between two windows is computed by dividing the intersection area of the two windows by the union area of the two windows, that is, intersection over union (IOU) (referred from [5]). The second is MABO vs. #WIN curve which shows the change of MABO with increasing the number of windows (proposals). With the same number of proposals, the higher the MABO, the better the method. The third one is detection rate (DR) vs. #WIN with a fixed IOU threshold. For each object, if the IOU between its groundtruth and a certain proposal is higher than a fixed IOU threshold, it can be considered as "detected" or "recalled". The detection rate is measured over all objects. When plotting this type of curve, we fix IOU threshold as 0.5. The last one is detection rate (DR) vs IOU threshold (Th) curve with the fixes number of windows. In the third and last metrics, the curves lies uppermost is the best one, that is, at the same x-axis value, the higher the y-axis value, the better the method.

When applying our proposed method to nighttime vehicle detection, we use the miss rate vs. false positives per image (FPPI) curve to evaluate the detection performance of some state-of-the art vehicle detection approaches. In this curve, at the same FPPI, the lower the miss rate, the better the method.

B. Independence, Weights and Performance of the Three Features

In our proposed Bayes saliency detection approach, we assume that the three features are independent with each other. In this paper, we use T-test to validate the independence of the three features. For each nighttime image, we extract the three features and compute the p-value obtained by T-test of each pair of the features (3 pairs). Then we compute the average p-value of each pair over all images. The results demonstrate that all average p-values of each pair are all less than 0.01, thus we can consider the three features are independent with each other.

When computing the likelihood functions of the observed salient pixels and the background, we design self-adapting weights of each feature for each class. The average weights over all images are demonstrated in Table I. The vehicle taillight map has the highest weights for both the background and the salient pixels, the local contrast is also important.

TABLE I
AVERAGE WEIGHTS AND PERFORMANCE OF EACH FEATURE

Feature	f_{lum}	f_{con}	f_{light}
Weight for the background	0.182	0.312	0.506
Weight for the salient pixels	0.160	0.321	0.519
MABO score	0.221	0.403	0.552

TABLE II
COMPARISONS OF MABO SCORE WHEN USING DIFFERENT SIZES OF THE FILTER

S_f	29×29	25×25	21×21	17×17	13×13
MABO score	0.8120	0.8213	0.8351	0.8262	0.8183
# proposal	19	17	15	16	17

The luminance has the lowest weights which means the luminance is less important and less effective than the two other features, and should be improved in the future. We also test the performance of our proposed approach when we only use one of the three features. The MABO score (computed using all proposals) of merely using vehicle taillight map (f_{light}) is higher than that of using two other features, which shows that the f_{light} should have higher weight.

C. Discussions on Parameters

In our proposed method, there are parameters which should be discussed. The size of the filter for extracting prior and the three features has impact on the map of prior and features, thus it might influence the performance of the saliency map and the final object proposals. We conduct experiments with different sizes of the filter and the comparisons are shown in Table II where "# proposal" denotes the largest number of extracted proposals of all images. The size used in this paper is $S_f = 21 \times 21$ which demonstrates the highest MABO when the fewest 15 proposals are computed. The MABO scores of various sizes are different but very close to each other, which suggests that the size of the filter has slight impact on the final proposals. This can be explained as follows: feature maps and prior maps are only low-level steps of our object proposal generator, their impacts are weakened by Bayes rule and our simple object proposal approach. By comparing the MABO scores of proposals generated merely using local contrast feature and vehicle taillight map, we find the window size ($S_w = 7 \times 7$) for computing Nakagami images during vehicle taillight map extraction and the size of the window ($S_w = 7 \times 7$) for computing local contrast demonstrate the highest MABO scores than other sizes. The multi-scale sliding windows are selected by observations and analysis of the sizes and saliency maps of the groundtruths. Thresholds of R_b are estimated via computing the R_b of the groundtruths and are precisely defined after extensive experiments that compare the performance of different thresholds. For the sizes of Gaussian filters used during score computing stage, we observe the scores of good windows and bad windows using different sizes. We find that when the sizes of $G_1(x, y)$ and $G_2(x, y)$ are $s_1 = 3/4\ L \times 3/4\ L$ and $s_2 = L \times L$, where L is the maximum between the height of window and the width of

TABLE III
THE SUMMARY OF PARAMETER MEANINGS AND SETTINGS

Parameter	Meaning	Value
S_f	Scale (mask size) of filter for computing prior and smoothing features	$S_f = 21 \times 21$
S_w	Window size for computing local contrast and Nakagami images	$S_w = 7 \times 7$
L_{th}, H_{th}	Low- and high-level thresholds on Nakagami images for detecting taillights	$L_{th} = 0.08, H_{th} = 11.6$
T_{opt}	Threshold for obtaining salient pixels and background	Computed using Eq. (2)
w_i^k	Weight of each feature for each class	Computed using Eq. (3)
s	Multi-scale window scale (width/height) for object proposal	$s \in \{20, 30, 40, ..., 200\}$
T_h	Threshold for segmenting background pixels from salient pixels	$T_h = 0.5$
R_b	Background pixel rate for deleting proposals	$0.05 < R_b < 0.5$
s_1, s_2	Filter sizes for computing scores of each proposal: small size(s_1)and large size(s_2)	$s_1 = 3/4L \times 3/4L, s_2 = L \times L$

window, the scores of good windows are all higher than that of bad windows. Table III summarizes the meanings and values of the parameters used in the proposed method.

D. Effectiveness of Bayes Saliency Detection

The accuracy of the Bayes saliency map has an impact on the performance of the final object proposals. To show the effectiveness and accuracy of the proposed Bayes saliency detection, we compared it with four general saliency detection methods: Context-aware saliency (CA) [45], Hierarchical Saliency Detection (HS) [46], the methods in [43] and [47], as well as some saliency detection approaches focusing on vehicles [48]–[50] in Fig. 4. As far as we know, most of current saliency detection methods focus on locating generic class of objects in daytime images such as the novel saliency propagation method in [51] and there is no work focusing only on nighttime vehicle saliency detection with RGB images. There are several daytime vehicle detection approaches based on saliency detection with RGB daytime images [48], [49], [52], far-infrared images [50] or vehicle detection in air-ground images [53]. In [48], the image saliency was detected using local features computed based on the color mutation and contrast between the block itself and its neighborhood blocks. Reference [49] utilized an approach to compute an image saliency map based on computing local saliency over random rectangular regions. Reference [52] used wavelet coefficients to reflect the level of uncommonness. Moreover, velocity information was employed to emphasize the mobile objects nearer to the driver. Reference [50] detected saliency pixels in far infrared images using features that were proposed in [54] for saliency detection in human detection. Reference [53] used a visual saliency step based on learned patch-based features by unsupervised learning to generate vehicle candidates for air-ground images. The optimal parameters of [48]–[50] were tuned in this work after our extensive experiments for fair comparison.

Our Bayes saliency map detects the two vehicles successfully despite the boundaries of the two vehicles are missed. The map of CA contains the vehicle on the right side but with much background, and divides the vehicle on the left side into two parts, which will lead to errors in the following proposal generation step. The map of HS contains the regions of the two vehicles but merges the two vehicles together, which makes it difficult to locate the two vehicles. The method in [43] mixes much background with the two vehicles and many road light regions are detected by [47]. The method in [48] highlights most pixels of the two vehicles but also the road light and lanes. In contrast, the method in [49] only highlights parts of the left vehicle and the road lamp but the right vehicle is not highlighted very well. The method in [50] highlights the vehicle lights but misses other parts of vehicles. From Fig. 4, we conclude that our Bayes saliency detection is better and more effective than other saliency detection approaches.

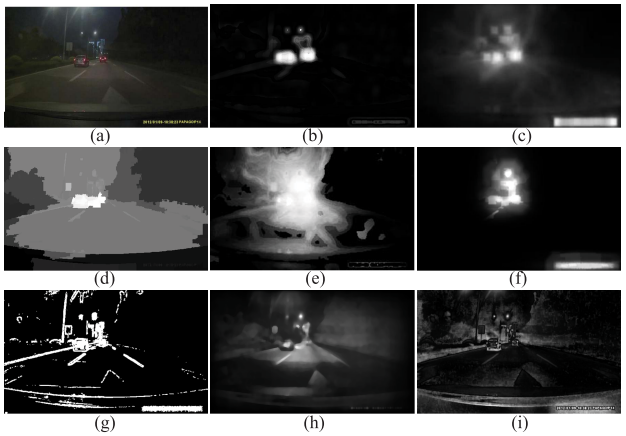


Fig. 4. Comparison between our proposed Bayes saliency detection and eight state-of-the-art saliency detection approaches. (a) Input nighttime image. (b) shows the result of our Bayes saliency map. (c), (d), (e) and (f) show results of general saliency detection approaches, i.e., CA [45], HS [46], [43] and [47] respectively. (g), (h), and (i) show results of saliency detection approaches focusing on vehicles, i.e., [48]–[50] respectively.

E. Object Proposal Results

In this section, we demonstrate the representative results of our object proposal generator in Fig. 5, where the green windows are the proposals that have high overlaps with the groundtruths of vehicles, and the red windows are the wrong proposals. From Fig. 5(a), our proposal generator can detect all salient vehicles in each image. The proposal results are closely related to the Bayes saliency map. The large bright regions in the saliency map are all incorporated in the proposal results. The number of proposals of our proposed approach is very sparse (less than 15), which can speed up the high-level tasks such as vehicle detection. In Fig. 5(b) a false positive (the red window that does not contain salient vehicles) is generated,

Fig. 5. Object proposal results of our proposed method. (a) Successful results; (b) Failure examples; (c) Results of headlights. The top row denotes the input nighttime images. The middle row represents the Bayes saliency map of each input image. The final object proposals of each image are shown in the bottom row. The correct and wrong detections are marked with the green and red rectangles, respectively.

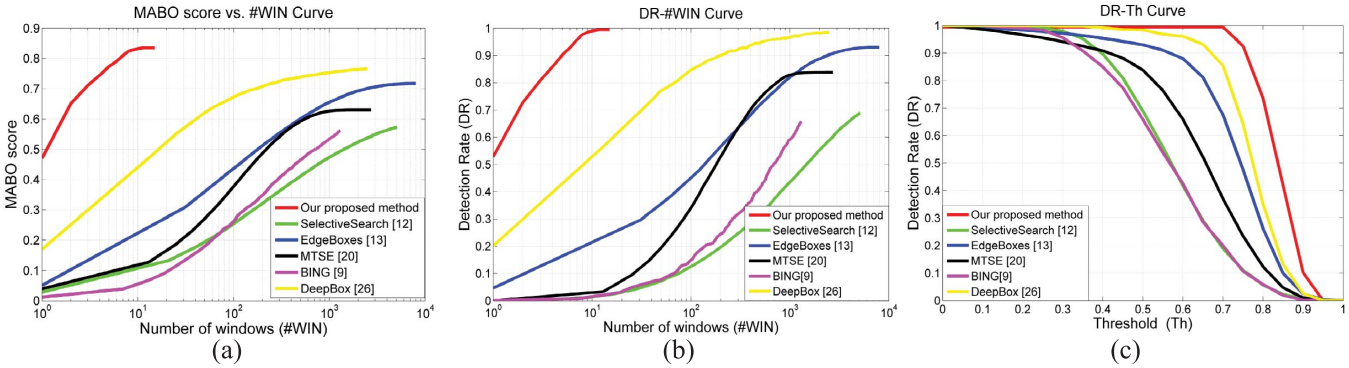


Fig. 6. (a)Tradeoff between the number of object proposals and the MABO score; (b) Tradeoff between # WIN and DR for our proposed approach and five state-of-the-art methods; (c) The comparison of DR-Th curves (Recall vs. IOU Threshold [9]) of each approach

because the saliency map considers the neon lights in this window to be salient. But this kind of false positives can be omitted when we use effective object detectors. We also find that small vehicles are missed, and one proposal might contain more than one small vehicle. From the saliency maps, we can conclude that these results are mainly caused by inaccurate saliency detection. In the future, we will design more accurate saliency detection method. We also show some results of detecting headlights in Fig. 5(c). Some oncoming vehicles with headlights are detected out but some are missed. Because the thresholding of our taillight detection is made to only guarantee the successful detection of taillights, which may be unsuitable for headlight detection. In the future, we will consider headlight detection and more accurate taillight detection.

F. Comparison With State-of-the-Art Methods

We utilize several evaluation metrics to assess the quality of our object proposals and compare our proposed method with the state-of-the-art methods. First, we compute the MABO scores [13] on the nighttime vehicle dataset developed in [3]. The nighttime dataset contains various classes of vehicle but we consider all classes of vehicle as the same class when we compute MABO. SelectiveSearch [13], EdgeBoxes [14] and MTSE [21] are directly used on the nighttime dataset. For BING [10] and DeepBox [26], we retrain the models on the training images of the nighttime dataset before evaluation. The comparisons of MABO scores are shown in Table IV, where "# proposal" denotes the largest number of proposals extracted using each method. The number of proposals obtained on some images might be smaller than "# proposal". In this case, we only use the generated proposals for evaluations. We find that our proposed method obtains the highest MABO score using the least number of proposals, which means that our proposed method generates a more accurate but smaller set of proposals than these state-of-the-arts for nighttime traffic images. Besides, we also compare these methods in terms of MABO score vs. #WIN curve [13] in Fig. 6(a). The curve of our proposed approach ends at about 15 proposals because the largest number of proposals generated by our proposed method is 15. Our proposed approach has the best good quantity/quality trade-off because our proposed approach reaches the highest MABO (0.8351) when only 15 proposals are generated.

Second, we also compare the computation time of our proposed object generator with some state-of-the-art methods in Table IV. To reduce the processing time, we resize the images to half of the orginal resolution of images. After

TABLE IV

COMPARISONS OF MABO SCORES AND COMPUTATION TIME OF EACH APPROACH

Approach	MABO score	# proposal	Time(s)
Our proposed method	0.8351	15	0.11
DeepBox [26]	0.7632	2500	6.54
SelectiveSearch [13]	0.5730	5100	3.21
BING [10]	0.5619	1300	0.02
EdgeBoxes [14]	0.7171	8000	0.41
MTSE [21]	0.6302	2730	0.26

generating object proposals, we resize the results to the original resolution for evaluation and other applications such as vehicle detection. The speed of our object proposal approach (0.11s per image) is only slower than BING [10] but obtains the highest MABO score with the smallest set of proposals.

We also evaluate DR-#WIN [10] of our proposed method and some state-of-the-arts on Sun Yat-sen University Nighttime Vehicle Dataset. DR-#WIN means detection rate (DR) given #WIN proposals. A vehicle is considered as being detected by a proposal if the strict PASCAL criterion, that is, the IOU [5] score is no less than 0.5, is satisfied. Fig. 6(b) shows the statistical comparison between our proposed method and state-of-the-art approaches: SelectiveSearch [13], BING [10], EdgeBoxes [14], MTSE [21], and DeepBox [26]. Our proposed approach demonstrates the higher DR than the five state-of-the-arts at any number of proposal windows and obtains about 0.9947 DR when about 15 proposals are generated because the IOU threshold is only 0.5 and our saliency map can detect almost all vehicles.

Finally, to demonstrate the impact of the overlap (IOU) threshold for the detection rate of each approach, we evaluate the DR of each approach under various thresholds and show the DR-Th curve (i.e., Recall vs. IOU Threshold curve [9]) in Fig. 6(c). Fig. 6(c) shows that our proposed approach demonstrates higher DR at the same threshold than the other four methods and still obtains high DR at a challenging threshold (0.8).

In summary, the extensive evaluations show that our proposed object proposal approach is more effective than several state-of-the-arts in terms of various evaluation metrics and the processing speed is fast (only 0.11s per image). DeepBox [26] is the next less performed than our proposed method in terms of proposal performance but it is the slowest.

We also compared some thresholding-based vehicle light detection approaches such as the HSV color threshold in [31], the multi-level thresholding in [32], the adaptive thresholding in [33] and the improved Ostu method in [35] with our proposed object proposal method. Considering that the vehicle candidates detected by their models are represented by the bounding boxes of vehicle taillights, which is different from our results outputting the boxes containing the whole vehicles, we use the detection rate as the metric for quantitative evaluation. For their methods, if the vehicle taillights of a vehicle are detected, this vehicle is detected accurately. For our proposed method, if a vehicle has at least one proposal that has high overlap (> 0.5), this vehicle is detected accurately.

TABLE V

THE COMPARISON BETWEEN OUR PROPOSED METHOD AND OTHER VEHICLE LIGHT DETECTION BASED METHODS

Method	Detection rate(%)
Our	99.47
The method in [31]	88.60
The method in [32]	86.20
The method in [33]	87.35
The method in [35]	88.40

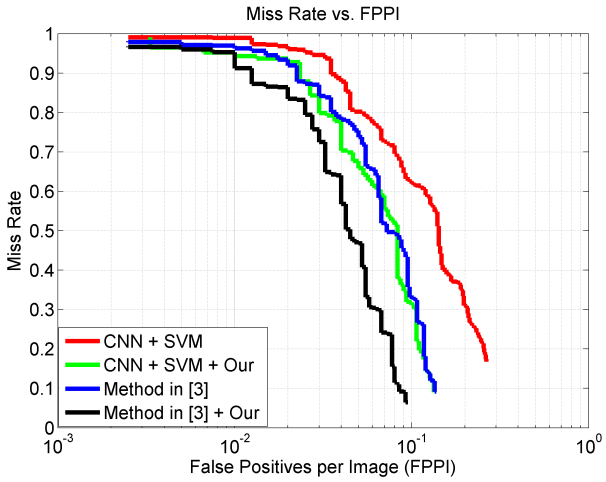


Fig. 7. The improvements of our proposed method for two state-of-the-art nighttime vehicle detection approaches.

The detection rate is the ratio of the number of vehicles detected accurately and the number of all vehicles. The comparison is shown in Table V. Our proposed method obtains the highest detection rate, which means our proposed method is better than these existing methods. This also indicates that generating vehicle candidates only using vehicle light detection might miss some vehicles.

G. Applications on Nighttime Vehicle Detection

Since object proposal generation is the key pre-processing step of object detection, we apply our proposed object proposal approach to nighttime vehicle detection to validate the effectiveness of our object proposals. We utilize two state-of-the-arts. In [3], regions of interest (ROIs) were extracted via weighted combining vehicle light detection with the scores of original EdgeBoxes [14]. The method in [3] extracted 30 ROIs. "CNN+SVM", that is, training support vector machine (SVM) classifiers using CNN features, is an effective framework of object detection [55]. We extract CNN features using Overfeat [56] from the top-30 proposals that are extracted by EdgeBoxes [14] and are with the highest scores for vehicle detection. We replace the object proposal approach using our proposed object proposal generator and keep the classification approaches unchanged. Fig. 7 compares the detection performance (miss rate vs. false positives per image (FPPI) curves) of two state-of-the-art nighttime vehicle detection approaches before and after using our proposed object proposal approach. We find that after using our proposed object proposal approach the detection performances of the two state-of-the-art methods are significantly improved, which validates

the effectiveness of our proposed object proposal approach for nighttime vehicle detection. Besides, we compare the detection time. Before using our proposed method, the detection speed of the method in [3] and CNN+SVM is 1.1s per image and 4.6s per image. Using our proposed method, they become both faster (the method in [3]: 0.28s per image and CNN+SVM: 1.6s per image).

Note that the edge map we used for computing the saliency prior is taken from [40], which trains a structured random forest for edge detection using daytime images (instead of nighttime images). Perhaps better results could have been obtained if the random forest is trained on nighttime imagery. However, we find that most edges in the vehicle regions of the Night-time Vehicle Dataset used here can be fast detected by the edge detection method in [40] trained from daytime images, although some edges of the vehicles in this nighttime dataset are missed because of the low lightness and contrast. Considering the quite high detection rate is finally obtained (i.e., 99.47%), we did not re-train the edge model in [40] on nighttime images. Nonetheless, it is a more rational strategy to train the random forest based edge detection model using nighttime images when we solve the task of nighttime vehicle detection.

V. Conclusions

This paper presents a novel object proposal generator for nighttime traffic images which combines the saliency map with object proposals. In the novel Bayes saliency detection approach, Bayes rule is used to integrate multiple features and compute the probability of a pixel belonging to salient pixels, and weights of each feature for each class are carefully designed. A new simple object proposal generator based on the obtained Bayes saliency map is adopted to generate a small set of proposals with high detection rate. Experimental results demonstrate the effectiveness of the proposed Bayes saliency detection and object proposal generator. Our proposed object proposal approach outperforms several state-of-the-art object proposal methods on a nighttime traffic image dataset in terms of various evaluation metrics. Our method obtains very high detection rate when only 15 proposals are generated. Our proposed object proposal approach can improve the detection performance and speed of the state-of-the-arts. In the future, we will apply deep learning into object proposals and focus on improving the existing object proposals to generate accurate object proposals on nighttime images.

References

[1] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2189–2202, Nov. 2012.

[2] N. An, J. Mittag, and H. Hartenstein, "Designing fail-safe and traffic efficient 802.11p-based rear-end collision avoidance," *Ad Hoc Netw.*, vol. 37, pp. 3–13, Feb. 2016.

[3] H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 57–65, May/Jun. 2016.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE CVPR*, Jun. 2009, pp. 248–255.

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.

[6] E. Rahtu, J. Kannala, and M. Blaschko, "Learning a category independent object detection cascade," in *Proc. IEEE ICCV*, Nov. 2011, pp. 1052–1059.

[7] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 222–234, Feb. 2014.

[8] J. Hosang, R. Benenson, and B. Schiele. (2014). "How good are detection proposals, really?" [Online]. Available: https://arxiv.org/abs/1406.6962

[9] N. Chavali, H. Agrawal, A. Mahendru, and D. Batra. (2015). "Object-proposal evaluation protocol is 'gameable.'" [Online]. Available: https://arxiv.org/abs/1505.05836

[10] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE CVPR*, Jun. 2014, pp. 3286–3293.

[11] J. Carreira and C. Sminchisescu, "CPMC: Automatic object segmentation using constrained parametric min-cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1312–1328, Jul. 2012.

[12] T. Lee, S. Fidler, and S. Dickinson, "Learning to combine mid-level cues for object proposal generation," in *Proc. IEEE ICCV*, Dec. 2015, pp. 1680–1688.

[13] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013.

[14] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. ECCV*, 2014, pp. 391–405.

[15] S. Manen, M. Guillaumin, and L. Van Gool, "Prime object proposals with randomized prim's algorithm," in *Proc. IEEE ICCV*, Dec. 2013, pp. 2536–2543.

[16] P. Rantalankila, J. Kannala, and E. Rahtu, "Generating object segmentation proposals using global and local search," in *Proc. IEEE CVPR*, Jun. 2014, pp. 2417–2424.

[17] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in *Proc. IEEE CVPR*, Jun. 2014, pp. 328–335.

[18] A. Humayun, F. Li, and J. Rehg, "RIGOR: Reusing inference in graph cuts for generating object regions," in *Proc. IEEE CVPR*, Jun. 2014, pp. 336–343.

[19] Y. Xiao, C. Lu, E. Tsougenis, Y. Lu, and C.-K. Tang, "Complexity-adaptive distance metric for object proposals generation," in *Proc. IEEE CVPR*, Jun. 2015, pp. 778–786.

[20] C. Wang, L. Zhao, S. Liang, L. Zhang, J. Jia, and Y. Wei, "Object proposal by multi-branch hierarchical segmentation," in *Proc. IEEE CVPR*, Jun. 2015, pp. 3873–3881.

[21] X. Chen, H. Ma, X. Wang, and Z. Zhao, "Improving object proposals with multi-thresholding straddling expansion," in *Proc. IEEE CVPR*, Jun. 2015, pp. 2587–2595.

[22] C. Lu, S. Liu, J. Jia, and C.-K. Tang, "Contour box: Rejecting object proposals without explicit closed contours," in *Proc. IEEE ICCV*, Dec. 2015, pp. 2021–2029.

[23] S. Liu, C. Lu, and J. Jia, "Box aggregation for proposal decimation: Last mile of object detection," in *Proc. IEEE ICCV*, Dec. 2015, pp. 2569–2577.

[24] J. Wang, J. Shen, and P. Li. (2015). "Object proposal with kernelized partial ranking." [Online]. Available: https://arxiv.org/abs/1502.01526

[25] N. Karianakis, T. J. Fuchs, and S. Soatto. (2015). "Boosting convolutional features for robust object proposals." [Online]. Available: https://arxiv.org/abs/1503.06350

[26] W. Kuo, B. Hariharan, and J. Malik, "DeepBox: Learning objectness with convolutional networks," in *Proc. IEEE ICCV*, Dec. 2015, pp. 2479–2487.

[27] A. Ghodrati, A. Diba, M. Pedersoli, T. Tuytelaars, and L. Van Gool, "DeepProposal: Hunting objects by cascading deep convolutional layers," in *Proc. IEEE ICCV*, Dec. 2015, pp. 2578–2586.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.

[29] J. H. Connell, B. W. Herta, S. Pankanti, H. Hess, and S. Pliefke, "A fast and robust intelligent headlight controller for vehicles," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Jun. 2011, pp. 703–708.

[30] A. López, J. Hilgenstock, A. Busse, R. Baldrich, F. Lumbreras, and J. Serrat, "Nighttime vehicle detection for intelligent headlight control," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2008, pp. 113–124.

[31] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010.

[32] Y.-L. Chen, "Nighttime vehicle light detection on a moving vehicle using image segmentation and analysis techniques," *WSEAS Trans. Comput*, vol. 8, no. 3, pp. 506–515, 2009.

[33] P. F. Alcantarilla *et al.*, "Automatic LightBeam controller for driver assistance," *Mach. Vis. Appl.*, vol. 22, no. 5, pp. 819–835, Sep. 2011.

[34] D. Jurić and S. Lončarić, "A method for on-road night-time vehicle headlight detection and tracking," in *Proc. Int. Conf. Connected Vehicles Expo (ICCVE)*, Nov. 2014, pp. 655–660.

[35] J. Wang, X. Sun, and J. Guo, "A region tracking-based vehicle detection algorithm in nighttime traffic scenes," *Sensors*, vol. 13, no. 12, pp. 16474–16493, 2013.

[36] Y. Li, N. Haas, and S. Pankanti, "Intelligent headlight control using learning-based approaches," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 722–727.

[37] S. Eum and H. G. Jung, "Enhancing light blob detection for intelligent headlight control using lane detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 1003–1011, Jun. 2013.

[38] C.-C. Wang, S.-S. Huang, and L.-C. Fu, "Driver assistance system for lane detection and vehicle recognition with night vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 3530–3535.

[39] R. K. Satzoda and M. M. Trivedi, "Looking at vehicles in the night: Detection and dynamics of rear lights," *IEEE Trans. Intell. Transp. Syst.* [Online]. Available: http://ieeexplore.ieee.org/document/7750549/

[40] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1558–1570, Aug. 2015.

[41] D. Y. Chen, Y. H. Lin, and Y. J. Peng, "Nighttime brake-light detection by Nakagami imaging," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1627–1637, Dec. 2012.

[42] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Mach. Learn.*, vol. 29, nos. 2–3, pp. 103–130, 1997.

[43] K.-F. Yang, H. Li, C.-Y. Li, and Y.-J. Li, "A unified framework for salient structure detection by contour-guided visual search," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3475–3488, Aug. 2016.

[44] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[45] 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.

[46] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE CVPR*, Jun. 2013, pp. 1155–1162.

[47] R. Margolin, L. Zelnik-Manor, and A. Tal, "Saliency for image manipulation," *Vis. Comput.*, vol. 29, no. 5, pp. 381–392, 2013.

[48] T. Wang, C. Xiu, and Y. Cheng, "Vehicle recognition based on saliency detection and color histogram," in *Proc. Chin. Control Decision Conf. (CCDC)*, May 2015, pp. 2532–2535.

[49] Y. Cai, H. Wang, X. Chen, L. Gao, and L. Chen, "Vehicle detection based on visual saliency and deep sparse convolution hierarchical model," *Chin. J. Mech. Eng.*, vol. 29, no. 4, pp. 765–772, Jul. 2016.

[50] Y. Cai, X. Sun, H. Wang, L. Chen, and H. Jiang, "Night-time vehicle detection algorithm based on visual saliency and deep learning," *J. Sensors*, vol. 2016, 2016, doi: 10.1155/2016/8046529.

[51] C. Gong *et al.*, "Saliency propagation from simple to difficult," in *Proc. IEEE CVPR*, Jun. 2015, pp. 2531–2539.

[52] X. Ma, X. Xie, J. Hu, and Y. Zhang, "Saliency analysis for car detection based on 2-D entropy and velocity prior," *Procedia-Social Behavioral Sci.*, vol. 138, pp. 378–385, Jul. 2014.

[53] Q. Xu, L. Jin, F. Jie, and S. Fei, "Air-ground vehicle detection using local feature learning and saliency region detection," in *Proc. 10th World Congr. Intell. Control Autom. (WCICA)*, Jul. 2012, pp. 4726–4731.

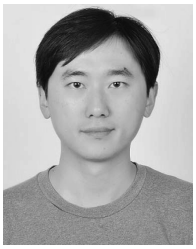
[54] S. Montabone and A. Soto, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," *Image Vis. Comput.*, vol. 28, no. 3, pp. 391–402, Mar. 2010.

[55] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE CVPR Workshops*, Jun. 2014, pp. 806–813.

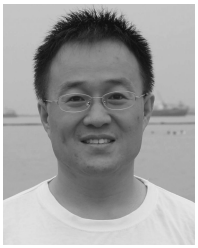
[56] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. (2013). "OverFeat: Integrated recognition, localization and detection using convolutional networks." [Online]. Available: https://arxiv.org/abs/1312.6229

Hulin Kuang received the B.Eng. and M.Eng. degrees from Wuhan University, China, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong. His research interests include computer vision and pattern recognition, especially object recognition.

Kai-Fu Yang received the Ph.D. degree in biomedical engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016. He is an Associate Professor with the Key Laboratory for Neuroinformation, Ministry of Education, School of Life Science and Technology, UESTC. His research interests include visual mechanism modeling, computer vision, and cognitive computing.



Long Chen (M'11) received the B.Sc. degree in communication engineering and the Ph.D. degree in signal and information processing from Wuhan University, Wuhan, China, in 2007 and 2013, respectively. From 2010 to 2012, he was a Ph.D. student co-supervised with National University of Singapore. He is currently an Associate Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His areas of interest include perception system of intelligent vehicle.



Yong-Jie Li (M'14–SM'17) received the Ph.D. degree in biomedical engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004. He is a Professor with the Key Laboratory for Neuroinformation, Ministry of Education, School of Life Science and Technology, UESTC. His research interests include visual mechanism modeling, image processing, and intelligent computation.

Leanne Lai Hang Chan (M'07) received the B.Eng. degree in electrical and electronic engineering from The University of Hong Kong and the M.S. degree in electronic engineering and the Ph.D. degree in biomedical engineering from University of Southern California. She is an Assistant Professor in electronic engineering with the City University of Hong Kong. Her research interests include artificial vision, retinal prosthesis, and neural recording.

Hong Yan (F'06) received the Ph.D. degree from Yale University. He was a Professor of Imaging Science with University of Sydney. He is currently a Professor of Computer Engineering with the City University of Hong Kong. He is a Fellow of IAPR. His research interests include image processing, pattern recognition, and bioinformatics.