Nighttime Vehicle Detection Based on Bio-Inspired Image Enhancement and Weighted Score-Level Feature Fusion

Hulin Kuang, Xianshi Zhang, Yong-Jie Li, *Member, IEEE*,
Leanne Lai Hang Chan, *Member, IEEE*, and Hong Yan, *Fellow, IEEE*

***Abstract*—This paper presents an effective nighttime vehicle detection system that combines a novel bioinspired image enhancement approach with a weighted feature fusion technique. Inspired by the retinal mechanism in natural visual processing, we develop a nighttime image enhancement method by modeling the adaptive feedback from horizontal cells and the center-surround antagonistic receptive fields of bipolar cells. Furthermore, we extract features based on the convolutional neural network, histogram of oriented gradient, and local binary pattern to train the classifiers with support vector machine. These features are fused by combining the score vectors of each feature with the learnt weights. During detection, we generate accurate regions of interest by combining vehicle taillight detection with object proposals. Experimental results demonstrate that the proposed bioinspired image enhancement method contributes well to vehicle detection. Our vehicle detection method demonstrates a 95.95% detection rate at 0.0575 false positives per image and outperforms some state-of-the-art techniques. Our proposed method can deal with various scenes including vehicles of different types and sizes and those with occlusions and in blurred zones. It can also detect vehicles at various locations and multiple vehicles.**

***Index Terms*—Object detection, feature extraction, high-level fusion, ROI extraction, image enhancement.**

I. Introduction

NIGHTTIME vehicle detection has become an important task in intelligent transportation systems (ITS) in recent decades. It is also one of the key technologies for advanced driver assistance systems (ADAS) and autonomous driving systems (ADS). About 30% of all vehicular accidents are caused by rear-end collisions that are one of the most fatal traffic accidents. In this paper, we focus on detecting moving vehicles in front of the driver at night to avoid rear-end collisions.

Some state-of-the-art object detectors are able to extract features from the original images taken in daytime [1], [2]. However, in nighttime images, the contrast between background and object, and the overall brightness are so low that some details of vehicles (e.g., edge, color, and shape features of vehicles) become unclear [3]. As a result, we should enhance the contrast, the brightness and details of the nighttime images before feature extraction for accurate vehicle detection. Inspired by the retinal information processing mechanisms of the biological visual system, we propose an effective nighttime image enhancement approach that models several important steps of the retinal information processing mechanism.

At night, the moving vehicles often turn on the taillights which are the most salient. Thus the taillights are very useful for extracting accurate regions of interest (ROIs). Generating a set of ROIs such as object proposal methods can improve the performance of current detection methods [4], [5]. In this paper, we adopt the ROI extraction approach proposed in [3] that combines vehicle taillight detection with EdgeBoxes [5].

Detection methods based on single features [1], [2], [6] have been proved to be effective. However, when dealing with more complex scenes, these types of detection methods might lead to misclassifications. Therefore, we extract not only features from convolutional neural network (CNN) [6], [7], but also compute two commonly used effective features: histogram of oriented gradient (HOG) [1] and local binary pattern (LBP) [8], to complement CNN features. Because we utilize multiple features, a key step is to combine them effectively. Score-level feature fusion has been reported to be effective [3], [9]–[12]. We focus on how to make full use of the complementarity of each feature and the different capabilities of the same feature for different classes, and then develop a score-level feature fusion approach that combines the three features with weights learnt from scores using a linear SVM.

The framework of the proposed nighttime vehicle detection approach is shown in Fig. 1. During the training stage, the original training samples are enhanced by the proposed image enhancement approach and then three complementary features: CNN features, HOG and LBP are extracted from the enhanced images. Three SVM classifiers are trained with LibSVM [13] and five-fold cross-validation is carried out using each individual feature respectively. The score vectors of each classifier

Manuscript received February 25, 2016; revised June 16, 2016; accepted July 21, 2016. Date of publication August 29, 2016; date of current version March 27, 2017. This work was supported in part by the City University of Hong Kong under Project 7002740, by the Major State Basic Research Program of China under Project 2013CB329401, and by the Natural Science Foundation of China under Project 91420105. The Associate Editor for this paper was S. S. Nedevschi. *(Corresponding authors: Leanne Lai Hang Chan; Yongjie Li.)*

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

X. Zhang and Y.-J. Li are with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: zhangxianshi@163.com; liyj@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.2016.2598192

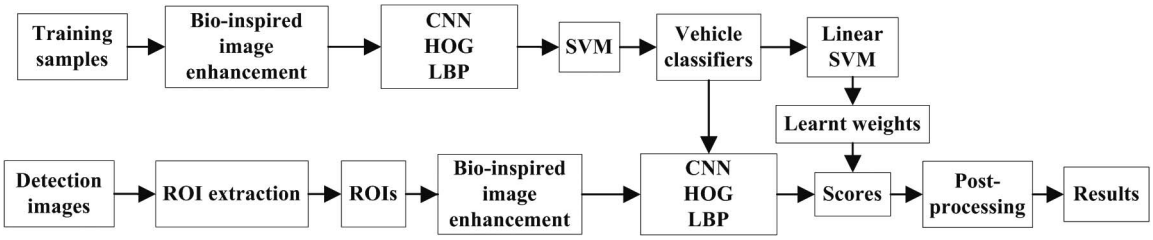


Fig. 1. Framework of the proposed nighttime vehicle detection method.

are also computed. The scores are used to learn weights of each feature for each individual class using a linear SVM (LibLinear [14]). During the detection stage, accurate ROIs are extracted from the input images. Subsequently, the three features are extracted from each enhanced ROI. Next, the trained classifiers are used to classify the corresponding features and compute scores which are summed with learnt weights to obtain final score vectors for prediction. Finally, post-processing (non-maximum suppression [2]) is performed to ensure that each vehicle is surrounded by a single window.

This paper makes the following contributions:

(1) A novel and effective bio-inspired nighttime image enhancement method is proposed to enhance the contrast, the brightness and details of nighttime images. Our approach is more effective than state-of-the-art methods.
(2) To complement CNN features, we extract HOG and LBP, and then utilize an effective weighted score-level feature fusion based on learnt weights using a linear SVM to combine the three features. The way to learn weights and bias terms is different from state-of-the-art classifier ensemble methods.

The rest of the paper is organized as follows. Related work is described in Section II. The details of the proposed bio-inspired image enhancement method are introduced in Section III. In Section IV, important steps of nighttime vehicle detection are described. The experimental results are illustrated in Section V. Finally, Section VI states the conclusions.

II. Related Work

A. Vehicle Detection

In recent decades, automatic vehicle detection techniques have been studied extensively [15]. Because there are so many vehicle detection methods, we only review several most relevant ones in this subsection. Some detection methods extract the features that can represent some attributes of the entire vehicle. Deformable parts model (DPM) extracted a feature that was considered as a variant of HOG for object detection [2]. Recently, each step of DPM was mapped to an equivalent CNN layer (Deep Pyramid DPM) [16] that outperformed some state-of-the-art CNN-based methods. Tehrani *et al.* improved the DPM method using latent filters for vehicle detection at night [17]. A simple framework where the extracted CNN features were used to learn a detector using SVM was proposed in [6] and it was proved to be very effective and efficient. The hybrid Deep Neuron Network (HDNN) was proposed in [18] to extract multi-scale features and this method was more effective than the traditional DNN for vehicle detection.

In nighttime scenes, the features of vehicle are not salient anymore. Thus, some vehicle detection methods focus on the vehicle lights, which are more salient [19]–[21] at night. In these methods, one or a pair of detected vehicle lights represents a vehicle. The highlighted head-lights were detected by a model-based detection method proposed in [20]. Light pixels were distinguished from reflection pixels by a Markov Random Field model that considered the reflection intensity and suppressed map, and the image intensity to represent and detect vehicles [21]. In [22], night vehicle recognition was achieved by searching taillights using image processing techniques. However, these methods are sensitive to scene variations because vehicle lights are difficult to detect accurately in complex scenes.

B. Nighttime Image Enhancement

The low contrast and brightness of nighttime images make nighttime vehicle detection challenging. Some simple nighttime image enhancement methods modified some traditional algorithms. For example, the classical histogram equalization algorithm was modified to maintain the color information of the nighttime images in [23]. Most of the nighttime image enhancement methods are image-fusion based [24], [25]. In these methods, the illuminant areas in nighttime images were enhanced by a variety of techniques to highlight the foreground areas. The final enhanced image was obtained by fusing background in day-time images and the highlighted foreground areas. Although effectiveness and robustness are reported, these are not appropriate for our case because we have no prior day-time images. Lin *et al.* [26] focused on improving the multi-scale retinex (MSR) for nighttime image enhancement without day-time images. They replaced the logarithm function with a customized sigmoid function. As a result, the output of MSR operation could represent an image directly. Due to fusing the original image with MSR image with proper weights, the noise can be suppressed and highlighted regions of original image can be preserved. Kuang *et al.* [3] improved the method in [26] by designing a new sigmoidal function, and weight function to make it work well in nighttime traffic scenes. The well-known MSR with color restoration (MSRCR) [27] was also used for low visibility images. After current nighttime image enhancement, noises might be magnified extensively, which leads to errors

in the next step of processing and insufficient local contrast enhancement.

C. Score-Level Feature Fusion

Concatenating one feature after another to obtain a new and large feature vector was a traditional and simple feature fusion approach (called "Concatenating" in this paper) [28] but it performed poorly when there was much redundant information among multiple features. Score-level fusion, that is, combining multiple features at score-level by fusing all scores obtained by each feature to acquire the final scores for classification prediction, has also been well studied [9]–[12]. Chen *et al.* utilized the mean, max, and product operators to combine two types of features at score stage [9]. Han *et al.* proposed a weighted feature fusion method based on Gaussian mixture models (GMM) [10]. The maximum likelihood probability was used as weights of each feature. Guermeur *et al.* combined two multi-class SVM classifiers with their proposed linear ensemble method (LEM) (called M-SVM+LEM in this paper) [11]. The classification results (i.e., scores) were post-processed and input into an LEM method to learn the weights of each classifier for each class but they did not learn the bias terms to calibrate the scores. Santos *et al.* applied a genetic algorithm (GA) to search for the optimal weights for each classifier to achieve fusion of the scores of each classifier (named GAFW in this paper) [12]. Kuang *et al.* computed the average classification contribution as weights of each feature to linearly combine the score vectors of features for nighttime vehicle detection (named "WSLF-ACC" in this paper) [3]. Although these methods have been reported to be effective, all weights should be computed again when adding new features or classifiers.

III. The Proposed Bio-Inspired Nighttime Image Enhancement Approach

In nighttime scenes, due to the low contrast and low brightness, the features of vehicles such as color and edge become unclear, which causes unsatisfactory performance of current object detection methods. For nighttime vehicle detection, nighttime image enhancement is a necessary pre-processing step. Effective and robust nighttime image enhancement approach should improve the contrast and the brightness of nighttime image, and also suppress noise (low brightness areas) and preserve highlighted areas that often contain some details of objects. However, current nighttime image enhancement methods cannot fulfill all the requirements mentioned above.

Inspired by the retinal information processing of the biological visual system, we propose a novel image enhancement approach whose framework is shown in Fig. 2. The details are introduced as follows.

Retinal information processing mechanism begins with the sampling by rod and cone photoreceptors. The red (R), green (G), and blue (B) components of the input image are processed respectively by long-, medium-, and short- wavelength cone photoreceptors of the retina [29].

Given an input image $I(x, y)$, it is first sampled by three types of cone photoreceptors, that is, converted to R, G, and B channel images (i.e., $I_c(x, y)$, $c \in (\mathrm{R}, \mathrm{G}, \mathrm{B})$). Then the horizontal cells (HCs) collect three cone signals from photoreceptors as

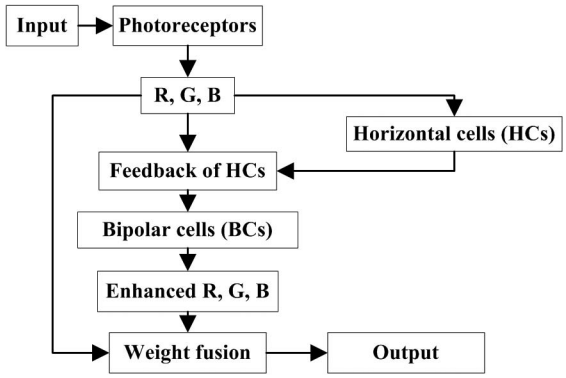


Fig. 2. Framework of the proposed nighttime image enhancement method.

$$\mathrm{HC}_{\mathrm{in}}(x, y) = \frac{1}{3} \sum_{c=1}^{3} I_c(x, y) \tag{1}$$

where $\mathrm{HC}_{\mathrm{in}}(x, y)$ is the brightness image of the input color image, which is normalized so that $\mathrm{HC}_{\mathrm{in}}(x, y) \in [0, 1]$.

HCs average the input from photoreceptors [30], [31]. This average function is influenced by the scale of the receptive field (RF) of HC and can be modeled as a Gaussian function. The scale of the RF (i.e., the standard deviation of Gaussian function) is modulated by the light [32], [33] and the local contrast [34]. Thus, the output of HC can be given by

$$\begin{aligned} \mathrm{HC}_{\mathrm{out}}(x, y) = {} & \omega_l \cdot \mathrm{HC}_{\mathrm{in}}(x, y) \otimes g\left(x, y, \sigma_l(x, y)\right) \\ & + \omega_c \cdot \mathrm{HC}_{\mathrm{in}}(x, y) \otimes g\left(x, y, \sigma_c(x, y)\right) \end{aligned} \tag{2}$$

where $\otimes$ denotes the convolution operator, $g(x, y, \sigma_l(x, y))$ and $g(x, y, \sigma_c(x, y))$ are two-dimensional Gaussian filters, $\sigma_l(x, y)$ and $\sigma_c(x, y)$ indicate the standard deviations (i.e., the scale of RF) decided by the light and the contrast respectively, ω_l and ω_c signify the weights of the average functions determined by the light and the contrast respectively. The light and local contrast are both very important. However, as far as we know, there is no report in literature to demonstrate whether the light is more or less important than the local contrast. Naturally, we experimentally set $\omega_l = \omega_c = 0.5$ in this work.

We simply use the brightness $\mathrm{HC}_{\mathrm{in}}(x, y)$ to estimate the light. The scale of RF of HCs is determined by the gap junctional coupling that is modulated by the light [32], [33]. Under dim starlight condition (low brightness), the conductance of the gap junctions is relatively low (small RF). As the ambient background light increases to twilight condition (intermediate brightness), the conductance increases (large RF). Under bright daylight condition (high brightness), the conductance is reduced again (small RF) [32], [33]. And the true relationship between the RF scale and light is similar to a continuous two-state Boltzmann function [34]. If we directly design $\sigma_l(x, y)$ as a continuous function simulating the true relationship, we should perform convolution operation using each possible RF scale, which is time-consuming. For reducing computation time

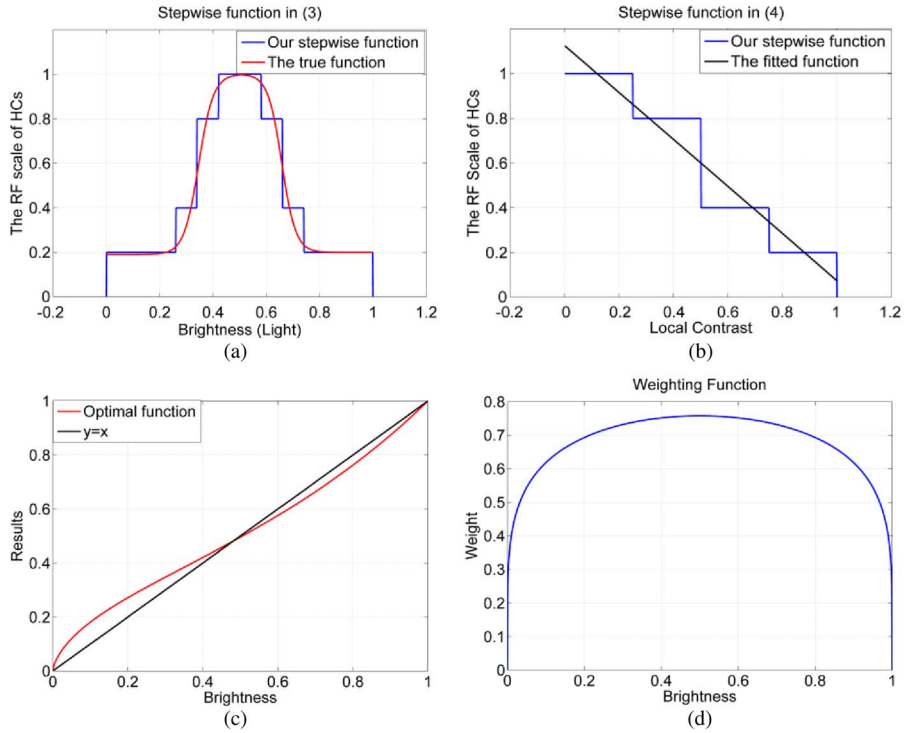


Fig. 3. Illustrations of several important functions used in our proposed image enhancement method. (a) The stepwise function in (3) and the true function between the light and the RF scale of HCs. (b) The stepwise function in (4) and the fitted function. (c) The shape of $\mathrm{HC}_{\mathrm{out}}(x,y)^{\lambda\mathrm{HC}_{\mathrm{out}}(x,y)+\lambda}$ with $\lambda = 0.65$. (d) The shape of the weighting function in (8) with $m = n = 0.2$.

we approximate the relationship between the RF scale and brightness as a stepwise function, which is written as

$$\sigma_l(x,y) = \begin{cases} \sigma/5 & \mathrm{HC}_{\mathrm{in}}(x,y) > u + 3s \\ 2\sigma/5 & u + 2s < \mathrm{HC}_{\mathrm{in}}(x,y) \le u + 3s \\ 4\sigma/5 & u + s < \mathrm{HC}_{\mathrm{in}}(x,y) \le u + 2s \\ \sigma & u - s < \mathrm{HC}_{\mathrm{in}}(x,y) \le u + s \\ 4\sigma/5 & u - 2s < \mathrm{HC}_{\mathrm{in}}(x,y) \le u - s \\ 2\sigma/5 & u - 3s < \mathrm{HC}_{\mathrm{in}}(x,y) \le u - 2s \\ \sigma/5 & \mathrm{HC}_{\mathrm{in}}(x,y) \le u - 3s \end{cases} \quad (3)$$

where u is the mean brightness of all pixels, s is the standard deviation of all pixels, σ is a predefined parameter that denotes the maximum scale of RF of HC and is set to be 1.0 (an empirical value based on extensive experiments) in this paper. The shape of function in (3) is shown in Fig. 3(a).

Besides, physiological findings indicate that the RF scale of HC at very low local contrast is at least twice larger than the scale at very large local contrast, and the larger the local contrast, the smaller the RF scale [35]. To accelerate the processing speed, this contrast-dependent mechanism is approximately modeled as a stepwise function given by

$$\sigma_c(x,y) = \begin{cases} \sigma/5 & C(x,y) > (uc + \max)/2 \\ 2\sigma/5 & uc < C(x,y) \le (uc + \max)/2 \\ 4\sigma/5 & (uc + \min)/2 < C(x,y) \le uc \\ \sigma & C(x,y) \le (uc + \min)/2 \end{cases} \quad (4)$$

where $C(x,y)$ is the local contrast of the brightness image, and uc, $\max$ and $\min$ denote the mean, maximum and minimum values of local contrast of all pixels, respectively. The local standard deviation of the pixel brightness within a 7 by 7 window is computed as local contrast in this paper. An illustration of function in (4) is shown in Fig. 3(b).

Then HCs send feedback signals to the photoreceptors to modulate the R, G, B cone signals and to locally regulate the image brightness within a normal range. $\mathrm{HC}_{\mathrm{out}}(x,y)$ is the Gaussian average of $\mathrm{HC}_{\mathrm{in}}(x,y)$, thus it varies in the range of [0, 1]. The result of dividing $I_c(x,y)$ by $\mathrm{HC}_{\mathrm{out}}(x,y)$ is larger than $I_c(x,y)$, which means this division can brighten the input image. Thus, we compute the modulated R, G, B cone signals, also the input of bipolar cells (BCs) as

$$\mathrm{BCin}_c(x,y) = \frac{I_c(x,y)}{a + \mathrm{HC}_{\mathrm{out}}(x,y)^{\lambda\mathrm{HC}_{\mathrm{out}}(x,y)+\lambda}} \quad (5)$$

where a is a predefined parameter to avoid dividing by zero and to control the enhancement degree. If a is large, the denominator of (5) might be larger than 1, which will decrease the brightness of image. If a is too small, the result of (5) might be much higher than $I_c(x,y)$, which will brighten the image extensively. After several trials with different a values, we set $a = 0.05$. In addition, we experimentally found that if we used $a + \mathrm{HC}_{\mathrm{out}}(x,y)$ directly as the denominator of (5), the noise (mostly with low brightness) would be over magnified, and meanwhile the bright pixels (e.g., with brightness > 0.5) would not be enhanced enough. To overcome this problem, we utilize $\mathrm{HC}_{\mathrm{out}}(x,y)^{\lambda\mathrm{HC}_{\mathrm{out}}(x,y)+\lambda}$ to increase

the denominator for low brightness pixels and decrease the denominator for the bright pixels. After comparing the shape of $\mathrm{HC}_{\mathrm{out}}(x,y)^{\lambda \mathrm{HC}_{\mathrm{out}}(x,y)+\lambda}$ with different λ values, we find that $\lambda = 0.65$ is a suitable setting (see Fig. 3(c)). With (5), the pixels of medium brightness become visually comfortable, and the low brightness pixels are magnified properly, without augmenting the noises too much.

The RF of most bipolar cells (BCs) consists of a smaller excitatory center and a larger inhibitory annular surround [36], which is commonly modeled by the "Difference of Gaussian" (DoG) model [37]. This center-surround interaction can reduce redundant information and improve spatial resolution and local contrast. Thus, the output of BCs can be generated as

$$\mathrm{BCout}_c(x,y) = \mathrm{BCin}_c(x,y) \otimes \left(g(x,y,\sigma_{cen}(x,y)) - k \cdot g(x,y,\sigma_{sur}(x,y))\right) \quad (6)$$

where $\sigma_{cen}(x,y)$ and $\sigma_{sur}(x,y)$ are the standard deviation of the excitatory center and its surrounding, respectively, and k controls the relative sensitivity of the inhibitory annular surrounding of RF. We set $\sigma_{sur}(x,y) = 2\sigma_{cen}(x,y)$ based on the physiological findings [36], [37]. In this work, after comparing the input and output of BCs with different parameters, we set $\sigma_{cen}(x,y)$ and k to be 0.5 and 0.2, respectively.

Now the brightness of the input image has been enhanced and the contrast has also been improved. However, by carefully observing the output of BCs, we found that there might be some noises (mostly with very low brightness) that would become very bright, and some highlighted areas containing object details would be enhanced so much that the details became blurred. Therefore, to suppress noise and preserve highlighted details, we combine the input image with the output of BCs to obtain the final enhanced image $F_c(x,y)$ as

$$F_c(x,y) = \omega(x,y)\cdot \mathrm{BCout}_c(x,y) + (1-\omega(x,y))\cdot I_c(x,y) \quad (7)$$

$$\omega(x,y) = \mathrm{HC}_{\mathrm{in}}(x,y)^m \cdot (1 - \mathrm{HC}_{\mathrm{in}}(x,y))^n \quad (8)$$

where $\omega(x,y)$ is the weighting function computed according to the brightness of each pixel and is used to decrease the final brightness of pixels whose original brightness is very low or high, and m and n control the shape of the weighting function. After observing the shape of weight functions with different parameters and comparing the images before and after weighting, we find that $m = n = 0.2$ is a suitable parameter setting that assigns low weights to the very low or high brightness pixels, thus can suppress noise and preserve highlighted areas in the input images (see Fig. 3(d)).

IV. Nighttime Vehicle Detection

A. ROI Extraction

We follow the ROI extraction framework in [3] but change the image enhancement to our proposed bio-inspired nighttime image enhancement method. First, the possible vehicle taillight regions are detected using the following steps: (1) converting the RGB images to intensity images and reducing noise using an empirical threshold (0.4 referred from [38]); (2) estimating the Nakagami images [38] utilizing a sliding window mechanism (9 by 9 window); (3) detecting possible vehicle taillight regions by two-stage thresholding where thresholds are decided by finding all manual located vehicle taillights and minimum non-taillight regions. Second, the coarse ROIs (each ROI have a score measuring its likeliness to be an object) are extracted by applying EdgeBoxes [5] on the enhanced images. Finally, a new score function that combines the coarse ROIs' scores and vehicle taillight detection together is constructed to calibrate scores and obtain more accurate ROIs. Similar to [3], we also select the top-30 windows as the final ROIs. The effectiveness of this ROI extraction approach has been validated in [3].

B. Weighted Score-Level Feature Fusion

Convolutional neural network (CNN) features have been reported to be very effective for object detection [6]. We extract CNN features using the codes from [39]. By observation, we find that some false positives caused by CNN features can be correctly detected when using LBP [8] or HOG [1]. Therefore, to complement CNN features, we extract LBP and HOG on grayscale images. In this paper, we linearly combine the three features using a score-level feature fusion approach where weight of each classifier for the individual class is learnt. Each feature demonstrates different accuracies for different classes, therefore learning the weight of each class respectively is more reasonable.

We first train three classifiers using each feature and SVM [13] respectively. The score vector of the ith sample using the jth classifier (i.e., feature) is demonstrated as

$$S_{ij} = [s_{i1j}, s_{i2j}] \quad (9)$$

where s_{i1j} and s_{i2j} denote the probability (score) of the ith sample to be classified as the background (class 1) and vehicle (class 2) using the jth classifier.

Inspired by "BING", an object proposal method [40] where the accurate and reasonable scores are computed by learning two terms using original scores as features, we learn a weight and a bias term of each classifier for each individual class.

The calibrated score vector of each classifier for each sample is obtained as

$$C_{ij} = w_j S_{ij} + b_j = [w_{j1}s_{i1j} + b_{j1}, w_{j2}s_{i2j} + b_{j2}] \quad (10)$$

where w_{j1} and w_{j2} are the learnt weights of the jth feature for background and vehicle respectively to control the importance of each score when computing the final scores, and b_{j1} and b_{j2} denote the bias terms of the jth feature for background and vehicle respectively to calibrate each score.

The terms w_{jk} and b_{jk} ($j \in \{1,2,3\}, k \in \{1,2\}$) are learnt by using a linear SVM (LibLinear [14]). All samples of the kth class are used as positive samples and the remaining samples are considered as negative samples. Scores of the jth classifier for the kth class are utilized as features and are input into a linear SVM. After we repeat these above steps 3 by 2 times, we can then learn the weight and bias term of each classifier for each class.

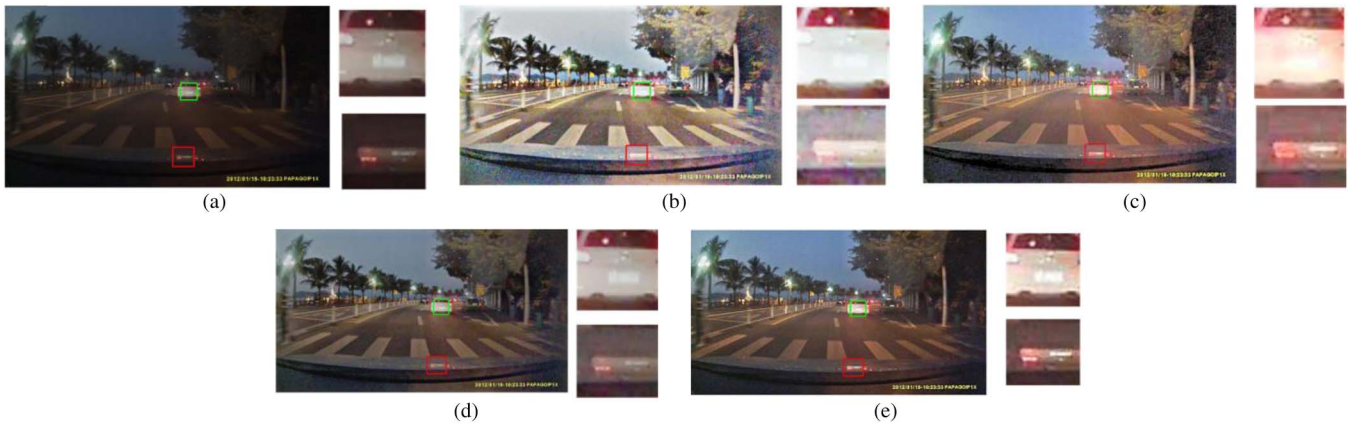


Fig. 4. Visual performances of our proposed image enhancement method and four other methods. (a) Original image. (b) The method in [26]. (c) MSRCR [27]. (d) The improved MSR in [3]. (e) Our proposed image enhancement method.

The final score vector of each sample is computed by summing all the calibrated score vectors linearly as

$$F_i = \sum_{j=1}^{3} C_{ij} = \left[\sum_{j=1}^{3} w_{j1} s_{i1j} + b_{j1}, \sum_{j=1}^{3} w_{j2} s_{i2j} + b_{j2} \right]. \quad (11)$$

During detection, for each ROI we first obtain the original score vector of each classifier. Then the final score vector is computed as (11) using the learnt weights and bias terms. The prediction class of each ROI is the index of the maximum value in the final score vector.

The benefits of our score-level feature fusion based on learnt weights are as follows:

(1) The strength of complementarity of each classifier (i.e., feature) for the individual class is fully utilized. For the same classifier, the weights of each class are different and learnt separately. Bias terms are learnt to calibrate the original scores. These aspects are neglected in some score-level feature fusion approaches.
(2) Our proposed feature fusion method can be easily used in multi-class detection tasks. Because the weight and bias term of each classifier for each class are learnt separately, when dealing with more features and more classes we only need to learn new weights and bias terms for new additional classes and features, and sum the new calibrated scores with the existing scores.
(3) Because each feature is used to train a classifier and the dimensionality of each feature is unchanged, our feature fusion approach can avoid the curse of dimensionality as well.

V. Experimental Results

A. Dataset

We evaluate our proposed nighttime vehicle detection approach on the nighttime vehicle dataset developed in [3]. This dataset includes 450 positive samples and 2000 negative samples in training set, 750 negative samples and 1000 positive samples in testing set, and 400 pictures (640∗360 pixels) containing 634 vehicles in detection set. These detection images are in various scenes like highways, housing estates, and bridges, which can validate the reliability and robustness of our method. The dataset is available online.[1]

B. Effectiveness of the Proposed Bio-Inspired Image Enhancement Method

To validate the effectiveness of the proposed bio-inspired image enhancement method, we compare it with the original image, the method in [26], MSRCR [27], and the improved MSR method in [3] on the detection set. First, we compare the visual performance of these approaches in Fig. 4 where the green window contains highlighted areas and details of the vehicle (e.g., plate region), and the red window includes some noise which has low brightness and some highlighted areas (e.g., red lights). The two windows are resized and shown at the right to demonstrate details of enhancement results. Although the method in [26] and MSRCR [27] can enhance the brightness of the whole image, the contrast is still low and they also magnify the noise so that the red bright areas and the plate region become blurred, which indicates that some details of vehicle are lost and highlighted areas are not preserved well. The improved MSR in [3] obtains higher contrast and can suppress noise and preserve highlighted regions, but the overall brightness of the image and the brightness of vehicle regions are not enhanced very well. Our proposed image enhancement approach can brighten the whole image, enhance the contrast between object and background, and suppress noise and preserve highlighted areas. Besides, the overall brightness is higher than the improved MSR in [3] and the brightness of vehicle region (i.e., the green window) is brightened and the details of vehicle are still clear. Based on visual inspection, our proposed image enhancement method exhibits the best performance.

We have also performed quantitative evaluation of our image enhancement method by computing some image quality assessment (IQA) metrics. For the full-reference (FR) IQA, because there is no reference image available in the dataset, we used the original detection images as the reference images to compute

[1] http://www.carlib.net

TABLE I
QUANTITATIVE EVALUATION OF OUR PROPOSED IMAGE ENHANCEMENT METHOD

Metric	Proposed method	Method in [3]	MSRCR [27]	Method in [26]	Original image
MSSIM	**0.832**	0.732	0.701	0.476	1.0
SSIM	**0.678**	0.612	0.448	0.311	1.0
NIQE	**2.986**	2.011	2.027	1.582	1.861
BRISQUE	**23.369**	17.993	25.182	13.566	15.936

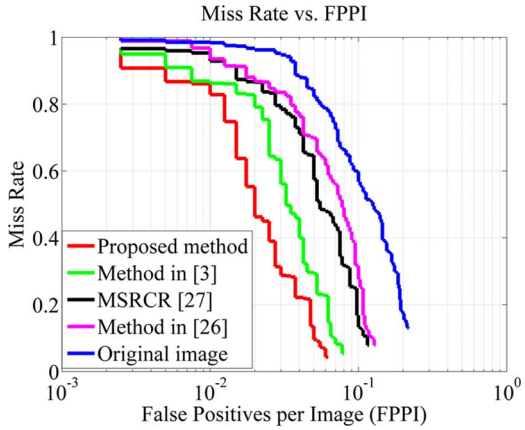


Fig. 5. Miss rate versus FPPI curves of different image enhancement methods.

two widely used full-reference IQA metrics: Multi-scale structural similarity (MSSIM) [41] and Structural similarity (SSIM) [42], which compute the similarity between enhanced images and reference images. We also utilized two commonly used no-reference (NR) IQA metrics: Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [43] that quantifies possible losses of "naturalness" and Naturalness Image Quality Evaluator (NIQE) [44] that analyze measurable deviations from statistical regularities observed in natural images. The quantitative evaluation is listed in Table I, where the metric values are the mean values over all the detection images. The higher the metric is, the better the method performs. We find that our image enhancement method obtains the highest MSSIM and SSIM, which indicates that our method can keep most useful information in the original images. The NIQE of our method is the highest and the BRISQUE of our method is only lower than that of MSRCR. These results demonstrate our image enhancement method indeed improves the image quality.

To validate the role of image enhancement in vehicle detection, we compared the miss rate vs. false positives per image (FPPI) curves in Fig. 5. To carry out this comparison, after conducting the image enhancement using these methods, we evaluated the proposed detection method on the detection set. In Fig. 5 the curve of our image enhancement approach is lower than all the other curves and demonstrates lower miss rate than the four other methods at the same FPPI. These results also validate that our bio-inspired image enhancement method is more effective than others for nighttime vehicle detection.

C. Validation of Feature Complementarity and Effectiveness of Our Weighted Feature Fusion

In this paper, to complement CNN features we extract HOG and LBP. We compared miss rate vs. FPPI curves of the proposed fusion and using each single feature. We validated the complementarity of the three features and the effectiveness of our weighted feature fusion in Fig. 6. In these comparisons, the differences are the features used, the other parts, i.e., image enhancement, ROI extraction and detection method are the same. From Fig. 6, CNN is better than HOG and LBP and the proposed fusion of CNN, HOG and LBP is better than using three single features. These results demonstrate that HOG and LBP are complements of CNN and our proposed fusion make full use of the complementarity of multiple features. From Table II and Fig. 6, we believe it is worth using a slightly longer processing time to achieve much better detection accuracy. We also compared our proposed feature fusion with some state-of-the-art methods in terms of the miss rate vs. FPPI curve to validate its effectiveness of our feature fusion method in Fig. 7. The only difference among these five methods is the feature fusion approach. We find that our feature fusion method is better than the other four feature fusion approaches. These results also demonstrate that the weights obtained by learning technologies are more reasonable and reliable.

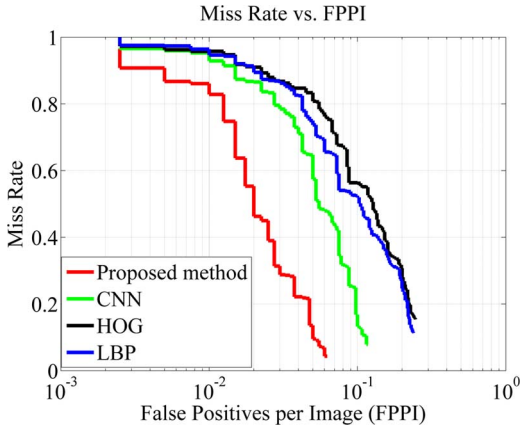


Fig. 6. Comparison between the proposed fusion of three features and the three single features.

TABLE II
COMPARISON OF DETECTION TIME AND FINAL DETECTION RATE

Method	Our	CNN [39]	HOG	LBP	Concatenating [28]
Time (sec per image)	**0.42**	0.31	0.21	0.22	0.58
Detection rate (%)	**95.95**	0.92	84.6	88.64	90.5
Method	DPM	CNN +SVM[6]	Method in [3]	Our +1/2 resolution	Our +1/4 resolution
Time (sec per image)	0.83	3.56	1.05	**0.23**	**0.14**
Detection rate (%)	73.7	91.5	93.34	**94.6**	**93.77**

D. Comparison With State-of-the-art Object Detection Methods

We select several state-of-the-art vehicle detection methods including DPM [2], convolutional neural networks features with

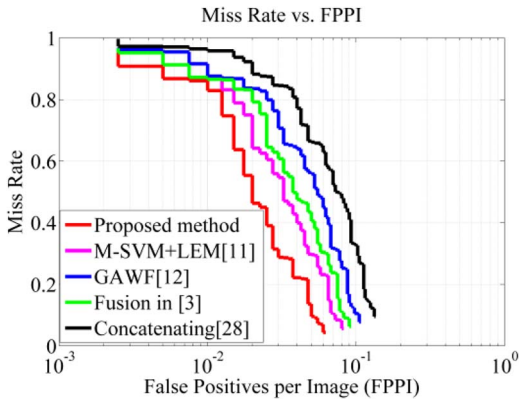


Fig. 7. Comparison between the proposed fusion and some state-of-the-art feature fusion approaches.

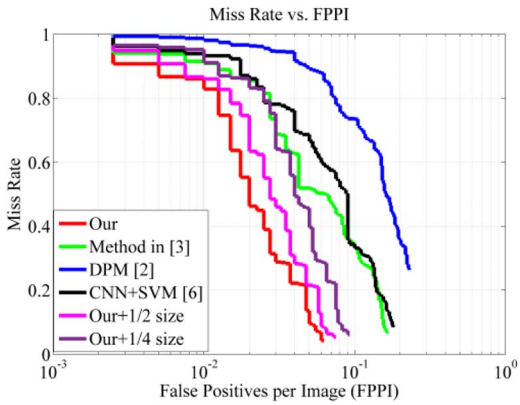


Fig. 8. Comparison between the proposed nighttime vehicle detection technique and some state-of-the-art detection methods.

SVM ("CNN+SVM") [6], and detection method in [3] for comparison. We trained a DPM classifier on the training set without image enhancement to detect vehicles in the detection images. For "CNN+SVM", we extracted the CNN features using Overfeat [7] on the training samples and used SVM to learn a detector. The method in [3] was developed on the same dataset, therefore we used the same curve reported in [3]. We also tested the detection performance of our method when we reduced the resolution of the detection image to half ("Our+1/2 resolution") or quarter ("Our+1/4 resolution"). After detection, the detection results were resized to the original resolution for evaluation. The comparisons are shown in Fig. 8 and Table II. Our proposed method ("Our") demonstrates the lowest miss rate at the same FPPI and the highest final detection rate (the right end point of curves), which shows our proposed method outperforms these state-of-the-art methods.

E. Detection Time

All experiments were conducted on a PC with Intel Core i7-4770 CPU @3.4GHz, 8GB RAM and GTX 970 graphics card. On average, the classifier training time for CNN, HOG and LBP is 58 mins, 4.4 mins and 5.1 mins (about 1 h 7 mins in total). Our method was trained offline before detecting

Fig. 9. Examples of detecting a single vehicle.

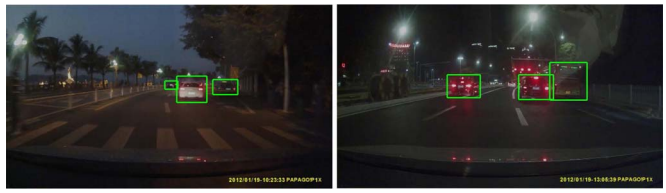


Fig. 10. Examples of detecting multiple vehicles within an image.

Fig. 11. Detection results in complex scenes.

vehicles. In this paper, we used the speedup framework in Fast R-CNN [39] to reduce the detection time. All ROIs were pooled into a feature map and then mapped to a feature vector by fully connected layers in CNN, thus we extracted features of 30 ROIs within an image simultaneously, which was faster than Overfeat [6], [7] that computed features of 30 ROIs one by one. Table II shows the detection time of some methods compared in this paper. The mean detection time of our model is about 0.42s per image including the time for image enhancement, ROI extraction, multi-feature extraction and feature fusion, vehicle detection, and post-processing.

We also tested whether reducing the resolution of the detection images can accelerate the detection process or not. We find that resizing the detection image to half or quarter of original resolution can reduce the processing time (see Table II). At the same time, the detection performance is degraded a little (Fig. 8), but still better than some state-of-the-arts. In the future, we will study how to improve the detection speed via simplifying the structure of CNN and developing faster ROI extraction methods.

F. Detection Results

Some examples of detection results by the proposed method are given in Figs. 9–14. The green rectangles denote the locations and sizes of the vehicles detected by our detection method. From Figs. 9–14, we find that our method successfully detects the vehicles in various numbers, types, locations and sizes from various scenes. Detection results in some complex scenes are shown in Fig. 11, where the vehicles with taillights on in front

Fig. 12. Detection results of blurred and partly occluded vehicles.

Fig. 13. Examples of missed vehicles.

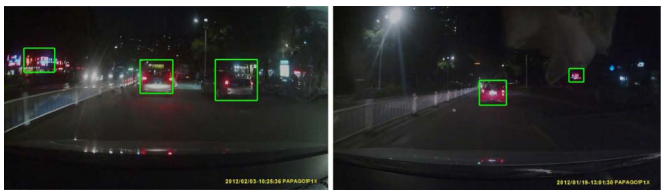


Fig. 14. Examples of false positives.

of the driver are all detected. Examples of detecting blurred and partly occluded vehicles are shown in Fig. 12. Blurred vehicles can be detected but the vehicles with most parts occluded by other vehicles are missed. In Fig. 13, we see most of the missed vehicles are small and far away from the driver. The distant small vehicles are less important for driving safety. Most of the false positives (Fig. 14) are caused by the bright areas that are not part of a vehicle but are similar to a vehicle to some extent. These results suggest that more accurate ROI extraction is necessary for robust nighttime vehicle detection.

VI. Conclusion

This paper proposes a novel nighttime vehicle detection method combining a novel bio-inspired image enhancement approach and a weighted score-level feature fusion strategy. Experimental results demonstrate that our nighttime image enhancement method can enhance contrast and brightness well and can also preserve and improve the object details, and is better than some state-of-the-art nighttime image enhancement techniques in terms of contributing to better vehicle detection. Moreover, three complementary features are combined linearly using a new weighted score-level feature fusion approach. This new feature fusion approach is more effective than using individual features and some state-of-the-art feature fusion approaches. Furthermore, our proposed vehicle detection method outperforms "DPM", "CNN+SVM" and a recent method in [3]. The vehicles in different types and sizes in complex scenes can be successfully detected. However, some partly occluded and distant vehicles are missed occasionally. In the future, we will focus on detecting the occluded and distant vehicles and accelerate the detection process using integral channel features in [45], [46]. In addition, we will try to build a larger and more complex nighttime vehicle dataset for validating and improving the vehicle detection system.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE CVPR*, 2005, vol. 1, pp. 886–893.

[2] 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.

[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] 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.

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

[6] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE CVPR*, 2014, pp. 512–519.

[7] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, and R. Fergus, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. ICLR*, 2014, pp. 1–16.

[8] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. ECCV*, 2004, pp. 469–481.

[9] Y. M. Chen and J. H. Chiang, "Face recognition using combined multiple feature extraction based on Fourier-Mellin approach for single example image per person," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 1833–1841, 2010.

[10] G. Han, C. Zhao, H. Zhang, and X. Yuan, "A new feature fusion method at decision level and its application," *Optoelectron. Lett.*, vol. 6, pp. 129–132, 2010.

[11] Y. Guermeur, "Combining multi-class SVMs with linear ensemble methods that estimate the class posterior probabilities," *Commun. Statist., Theory Methods*, vol. 42, no. 16, pp. 3011–3030, 2013.

[12] S. Chernbumroong, S. Cang, and H. Yu, "Genetic algorithm-based classifiers fusion for multisensor activity recognition of elderly people," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 282–289, Jan. 2014.

[13] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 389–396, 2011.

[14] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.

[15] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle detection techniques for collision avoidance systems: A review" *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 1–21, Oct. 2015.

[16] R. Girshick, F. Iandola, T. Darrell, and J. Malik, "Deformable part models are convolutional neural networks," in *Proc. CVPR*, 2014, pp. 437–446.

[17] H. Tehrani, T. Kawano, and S. Mita, "Car detection at night using latent filters," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 839–844.

[18] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[19] T. Schamm, C. V. Carlowitz, and J. M. Zollner, "On-road vehicle detection during dusk and at night," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 418–423.

[20] K. Robert, "Video-based traffic monitoring at day and night vehicle features detection tracking," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2009, pp. 1–6.

[21] W. Zhang, Q. M. J. Wu, G. Wang, and X. You, "Tracking and pairing vehicle headlight in night scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 140–153, Mar. 2010.

[22] K. B. K. Hemanth, A. Naik, and A. Gowda, "Vehicle recognition at night based on tail light detection using image processing," *Int. J. Res. Eng. Sci.*, vol. 2, no. 5, pp. 68–75, 2014.

[23] M. S. Sayed and J. Delva, "Low complexity contrast enhancement algorithm for nighttime visual surveillance," in *Proc. Int. Conf. Intell. Syst. Des. Appl.*, 2010, pp. 835–838.

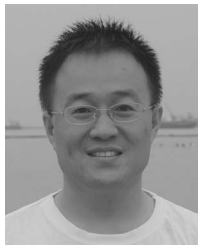
[24] Pallavi and R. Sharma, "A novel algorithm for nighttime context enhancement," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 7, pp. 444–447, 2013.
[25] Y. Rao and L. Chen, "A survey of video enhancement techniques," *J. Inf. Hiding Multimedia Signal Process.*, vol. 3, no. 1, pp. 71–99, 2012.
[26] H. Lin and Z. Shi, "Multi-scale retinex improvement for nighttime image enhancement," *Optik*, vol. 125, no. 24, pp. 7143–7148, 2014.
[27] Z. U. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *J. Electron. Imag.*, vol. 13, no. 1, pp. 100–110, 2004.
[28] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. IEEE ICCV*, 2009, pp. 32–39.
[29] R. H. Masland, "The neuronal organization of the retina," *Neuron*, vol. 76, pp. 266–280, 2012.
[30] C. Joselevitch, "Human retinal circuitry and physiology," *Psychol. Neurosci.*, vol. 1, pp. 141–165, 2008.
[31] M. Kamermans, I. Fahrenfort, K. Schultz, U. Janssen-Bienhold, T. Sjoerdsma, and R. Weiler, "Hemichannel-mediated inhibition in the outer retina," *Science*, vol. 292, pp. 1178–1180, 2001.
[32] D. Xin and S. A. Bloomfield, "Dark- and light-induced changes in coupling between horizontal cells in mammalian retina," *J. Comput. Neurol.*, vol. 405, pp. 75–87, 1999.
[33] S. A. Bloomfield and B. Völgyi, "The diverse functional roles and regulation of neuronal gap junctions in the retina," *Nature Rev. Neurosci.*, vol. 10, pp. 495–506, 2009.
[34] M. Srinivas, M. Costa, Y. Gao, A. Fort, G. I. Fishman, and D. C. Spray, "Voltage dependence of macroscopic and unitary currents of gap junction channels formed by mouse connexin50 expressed in rat neuroblastoma cells," *J. Physiol.*, vol. 517, pp. 673–689, 1999.
[35] X.-M. Song and C.-Y. Li, "Contrast-dependent and contrast-independent spatial summation of primary visual cortical neurons of the cat," *Cerebral Cortex*, vol. 18, pp. 331–336, 2008.
[36] A. Kaneko and M. Tachibana, "Double color-opponent receptive fields of carp bipolar cells," *Vis. Res.*, vol. 23, pp. 381–388, 1983.
[37] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *J. Physiol.*, vol. 187, pp. 517–552, 1966.
[38] 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.
[39] R. Girshick, "Fast R-CNN," in *Proc. IEEE ICCV*, 2015, pp. 1440–1448.
[40] M. M. Cheng, Z. Zhang, W. Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE CVPR*, 2014, pp. 3286–3293.
[41] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.
[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[43] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[44] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 209–212, Mar. 2013.
[45] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. BMVC*, 2009, pp. 91.1–91.11.
[46] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second," in *Proc. IEEE CVPR*, 2012, pp. 2903–2910.

Hulin Kuang received the B.Eng. and M.Eng. degrees from Wuhan University, Wuhan, 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, Kowloon, Hong Kong.

His current research interests include computer vision and pattern recognition, particularly object detection.

Xianshi Zhang received the M.Sc. degree in automation engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2008, where he is currently working toward the Ph.D. degree in biomedical engineering.

His research interests include visual mechanism modeling and image processing.



Yong-Jie Li (M'14) received the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004. He is currently 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'11) received the B.Eng. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong and the M.S. degree in electrical engineering and the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles, CA, USA.

She is currently an Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, 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, New Haven, CT, USA.

He was a Professor of imaging science with The University of Sydney, Sydney, Australia, and is currently a Professor of computer engineering with the City University of Hong Kong, Kowloon, Hong Kong. His research interests include image processing, pattern recognition, and bioinformatics.

Dr. Yan is a Fellow of the International Association for Pattern Recognition.