

Received 28 May 2023, accepted 17 June 2023, date of publication 26 June 2023, date of current version 30 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289595



# Multi-Exposure Fusion With Guidance Information: Night Color Image Enhancement for Roadside Units

CHEN YAQING<sup>ID</sup>, (Member, IEEE), AND WANG HUAMING<sup>ID</sup>

College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Jiangsu 210016, China

Corresponding author: Wang Huaming (hmwang@nuaa.edu.cn)

**ABSTRACT** The computer vision-based roadside occupation surveillance system is a key infrastructure component of Cooperative Intelligent Transport Systems. However, traffic images captured under low-light conditions suffer from low visibility and unexpected noise. Despite the great progress achieved in recent years, the existing night image enhancement algorithms often suffer from color deviation, ghosting, and overexposure problems in practical traffic applications. Thus, we present a novel night color image enhancement approach to overcome this issue by combining multi-sensor fusion and pseudo-multi-exposure fusion techniques. Unlike the traditional exposure adjustment-based approaches, we performed a novel bidirectional region segmentation-based inverse tone mapping operator to generate pseudo-multi-exposure sequences from day and night image pairs. Meanwhile, to solve the problem that moving objects are diluted after fusion, a partial differential equation (PDE)-based luminance stretching is applied to the moving areas to guarantee that the enhanced image always highlights the traffic targets. Instead of image feature-based methods for moving object detection, we generate more accurate moving regions by fusing data from the radar and camera sensors. Finally, a pyramid-based fusion method with an improved weight function is conducted to generate high-quality traffic images. The proposed method and five state-of-the-art methods are evaluated on randomly selected images from the Rope-3D database and nighttime images captured by an Intelligent Roadside Surveillance System. The experimental results demonstrate that our method has significant advantages in enhancing details and making colors more natural for human observation.

**INDEX TERMS** Intelligent transport, roadside units, image enhancement, multi-exposure fusion.

## I. INTRODUCTION

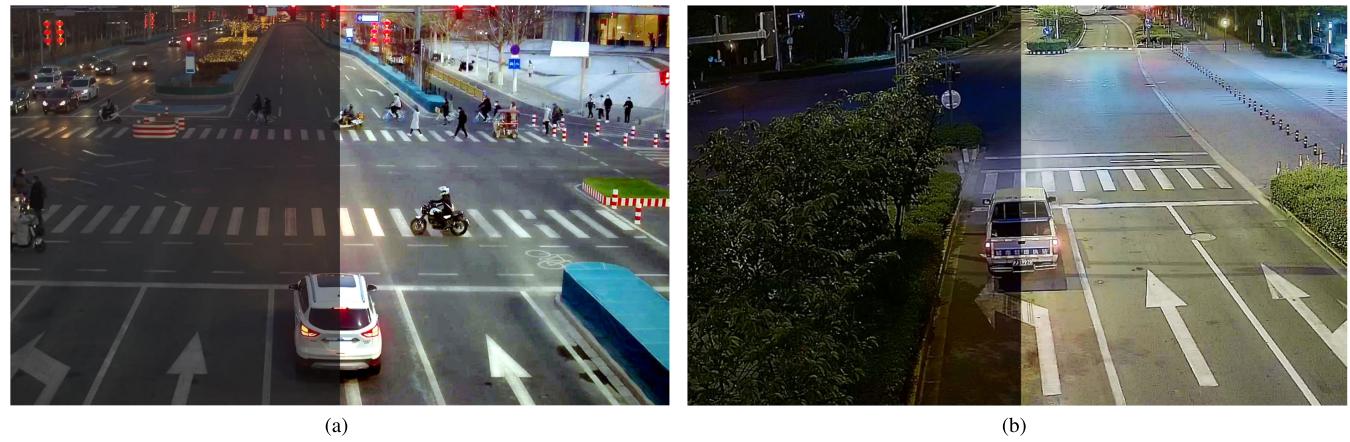
Cooperative Intelligent Transport Systems (Cooperative ITS or C-ITS), also known as vehicle-road cooperation systems and vehicle-road collaboration systems, enable cooperation and communication between autonomous vehicles, intelligent infrastructure, and traffic control centers. C-ITS has been the driving force to enable autonomous vehicles, smart cities, and the Internet of Things (IoT) [1]. As one of the most valued technologies in C-ITS, the Computer Vision-Based roadside occupation surveillance system (Roadside units or RSU) provides extended coverage and more traffic information dimensions than vehicle sensors. With the flexible

The associate editor coordinating the review of this manuscript and approving it for publication was Yanli Xu<sup>ID</sup>.

deployment and ease of operation, RSU has successfully been applied in the C-ITS [2].

However, Traffic images captured under low-light conditions suffer from low visibility and unexpected noise, and therefore Vision-Based roadside units often work inefficiently or even fail to operate at night. All these phenomena urge a vigorous need for solutions to address image quality degradation problems.

Although some existing image enhancement algorithms, such as histogram enhancement, Retinex algorithm, and wavelet decomposition, have achieved good results in data sets and some specific scenarios [3], [4], and [5], these are often inefficient and unstable in practical traffic application. One of the most important reasons is that the degradation of night images captured by low-cost traffic cameras



**FIGURE 1.** Some examples from the proposed method (left part: original, right part: enhanced).

is sometimes too serious about meeting the preconditions for the above single-frame image enhancement. Some deep learning-based image enhancement methods can achieve better performance after being properly trained, such as Dual-Channel Dehazin-Net [6]. However, obtaining proper training data is challenging. To overcome the above issues, context enhancement-based techniques, such as multi-sensor fusion (MSF) and multi-exposure fusion (MEF), have been widely used in C-ITS. Using various dimensions of information can improve the quality of the degraded images captured under poor illumination or nighttime, thus meeting the requirements of intelligent applications [7], [8]. Some MEF-based methods can merge differently exposed low dynamic range (LDR) images to generate a high dynamic range (HDR) image. However, it is not easy to obtain multiple images in traffic scenes. Thus, the proposed method generates pseudo-multi-exposure sequences from day and night image pairs to overcome this issue.

This paper proposes a novel pseudo-multi-exposure fusion-based image enhancement algorithm for low-light traffic images via multi-source data fusion. By the decision level fusion of camera and radar and pixel level fusion of day and night images, our method can improve the quality of the nighttime images and significantly enhance the texture of key traffic participants such as vehicles and pedestrians.

The suggested method is extensively evaluated on the Rope3D dataset and nighttime images captured by an Intelligent Roadside Surveillance System, demonstrating the effectiveness and generalization of our approach. Fig. 1 depicts some examples from the proposed method.

This paper's contributions are as follows:

- A novel night color image enhancement approach is introduced by combining multi-sensor and pseudo-multi-exposure fusion techniques.
- A region-based tone mapping method is proposed to generate the multi-exposure sequence from day and night image pairs.
- The data from the radar sensors are exploited to generate moving regions in the image, and a PDE-based

luminance stretching is further applied to these areas to highlight the traffic targets.

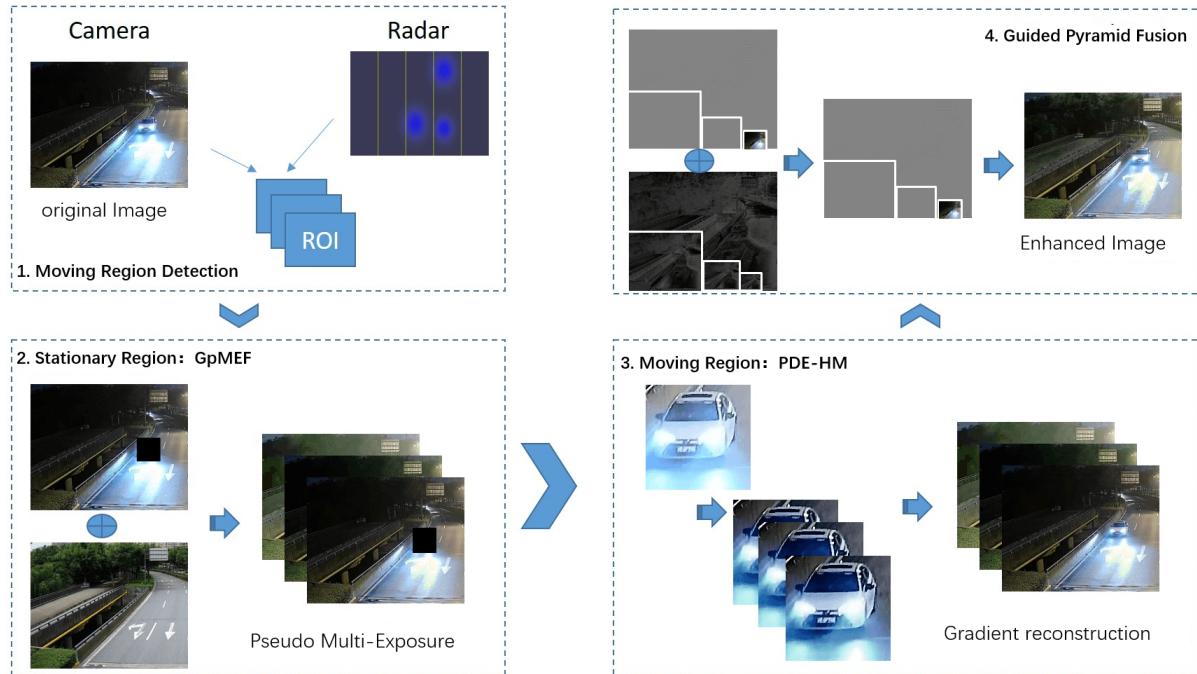
- An improved weighting function for the pyramid fusion is used to generate high-quality traffic images. Specifically, we consider guidance information in the weight calculation process.

The remainder of this paper is organized as follows. Section II reviews the related works about MEF. Section III proposes a novel Pseudo Multi-Exposure Fusion-based image enhancement algorithm for traffic low-light images via multi-source data fusion. Section IV discusses the experimental results, and finally, Section V concludes this paper and suggests some future research directions.

## II. RELATED WORKS

Recently, many Multi-exposure fusion (MEF) methods have been applied to intelligent transportation systems for their excellent capability of context enhancement [9], [10]. Compared to the single image enhancement method, MEF can better use the advantages of intelligent traffic scenes, such as easy-to-carry-out multi-source sensor fusion and relatively fixed vision. The existing MEF techniques can be categorized into four categories [11], [12], and [13]: spatial domain methods, transform domain methods, the combination of spatial domain and transform domain methods and deep learning methods [14], [15] [16], and [17].

Two key steps of MEF are image sequence generation and feature fusion, with different MEF methods relying on different techniques to obtain suitable source image sequences. The most common approach is acquiring multiple Low Dynamic Range (LDR) images of the same scene under different exposure. When obtaining multiple images with the cameras becomes challenging, the pseudo-exposure MEF is the most appropriate method, as it can generate pseudo-multi-exposure image sequences using only one or several LDR images [18]. For instance, Kinoshita et al. [19] generated pseudo-exposure image sequences imitating different exposure levels by simulating the camera exposure-response



**FIGURE 2.** General framework of the proposed method.

function and then performed fusion enhancement. Paper [20] divided the image into regions and used different S-curves for tone transformation in different regions, thereby generating pseudo-exposure image sequences that enhanced the details of different regions. Wang. et al. brighten the low-light image by fusing the image sequence generated by CNN [21]. It should be noted that this type of technology does not require additional cameras or other sensors, significantly reducing the complexity of the entire system.

Feature fusion of multiple images is another key step of MEF. The traditional method is a weighted fusion [14]. Although this method is simple and efficient, an elegant weighting function requires expert knowledge and experience, heavily depending on laborious attempts and mistakes. With the development of information fusion technology in recent years, multi-scale feature fusion technology has received more extensive attention. Typically, multi-scale features are normalized through upsampling and downsampling operations to achieve pyramid fusion, such as PGA-Net [22] and DenserNet [23]. An alternative way is integrating multi-scale features [24]. For example, Zhang et al. designed a fusion cell based on Neural Architecture Search, allowing the network to search for the best multi-scale fusion methods [25]. After full training, these methods have achieved appealing results in multiple public data sets, but in practical applications, a large number of scene data must be collected for migration training to achieve stable results in specific tasks. The idea based on multi-scale feature pyramid fusion is still the main research direction of MEF. In 2009, Mertens et al. first proposed a multi-exposure image fusion algorithm based on an image pyramid [11]. Then, a large number of

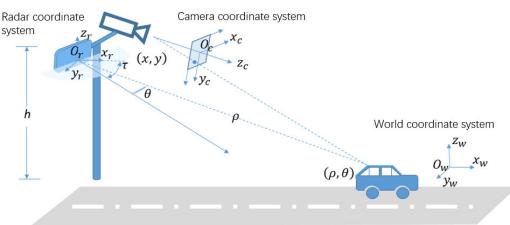
MEF algorithms have been proposed based on improved Laplace Pyramid (LP), such as edge-preserving smoothing pyramids. Nevertheless, although it has achieved certain results, its fusion process mostly relies on statistical index evaluation, and it is difficult to add guidance information such as traffic scene prior knowledge. This limitation urges us to use guided multi-level semantic features to improve the fusion effect.

### III. PROPOSED ALGORITHM

Fig. 2 illustrates the general framework of the proposed method, where the four boxes contain this study's key steps. (i) Moving Region Detection: the original image is divided into moving and stationary regions using radar detection information. (ii) Guided pseudo-multi-exposure fusion (GpMEF): a region-based pseudo-multi-exposure fusion is proposed based on the idea that the mapping relationship between low light and reference images obtains the exposure sequences. (iii) PDE histogram modification (PDE-HM): we propose a histogram-based partial differential equation to synchronize contrast enhancement and gradient-domain reconstruction. (iv) Guided Pyramid Image Fusion (GPIF): the exposure fusion method based on the Laplacian pyramid is utilized to fuse the pseudo-exposure image sequence by using the reference image as guidance information.

#### A. MOVING REGION DETECTION

Despite many ghosting-free exposure fusion methods proposed recently, which have achieved high-quality results, these are limited to daytime for it is difficult to point

**FIGURE 3.** Coordinate system setups.

out moving vehicles and pedestrians from low-quality images [26].

One possible solution is the sensor fusion technique [27]. Hence, this paper exploits the data from the radar sensors to generate a region containing moving objects (Moving Region) in the video image. Specialized image enhancement strategies are used in different regions.

Fig. 3 illustrates the coordinate systems required to map the relationship between the radar and camera detections:

### 1) RADAR COORDINATE SYSTEM

In the polar coordinate system, having the radar position as the origin, the target state vector is expressed as  $X_{radar} = [\rho \theta S_r v]$ , where  $\rho$  is the distance,  $\theta$  is the angle,  $S_r$  is the cross-sectional reflection radius, and  $v$  is the velocity.

### 2) CAMERA COORDINATE SYSTEM

In the camera coordinate system, the target region is expressed as  $X_{cam} = [x_c \ y_c \ W_{roi} \ H_{roi}]$ , where  $(x_c, y_c)$  are the target pixel coordinates, and  $(W_{roi}, H_{roi})$  are the width and height of the moving region, calculated based on the following equation.

$$\begin{cases} W_{roi} = k \cdot (\max\{x_{c,\partial\Omega}\} - \min\{x_{c,\partial\Omega}\}) \\ H_{roi} = k \cdot (\max\{y_{c,\partial\Omega}\} - \min\{y_{c,\partial\Omega}\}) \end{cases} \quad (1)$$

where  $P_{c,\partial\Omega} = (x_{c,\partial\Omega}, y_{c,\partial\Omega})$  are the pixel coordinates of the corner points of the 3D bounding boxes of traffic objects, expressed as (2).  $k$  is the magnification factor of the ROI area, set as 1.2 to ensure that the selected area contains both the moving target and as little background as possible.

$$\begin{cases} P_{c,\partial\Omega}^{bottom} = R_t \cdot P_{r,\partial\Omega} \\ P_{c,\partial\Omega}^{top} = R_t \cdot fheight(P_{r,\partial\Omega}, h) \\ fheight(P, h) = \frac{H}{H-h}P \end{cases} \quad (2)$$

where  $P_{c,\partial\Omega}^{bottom}$  are the four bottom corner points of the 3D bounding box, while  $P_{c,\partial\Omega}^{top}$  are the four top corner points.  $fheight(P, h)$  calculates the equivalent position deviation generated by the target height, and  $H$  and  $h$  are the heights of the camera and the traffic object. The  $R_t$  represents the translational perspective conversion matrix generated by the installation position of the radar, calculated by the Direct Linear Transformation method (DLT) [28].

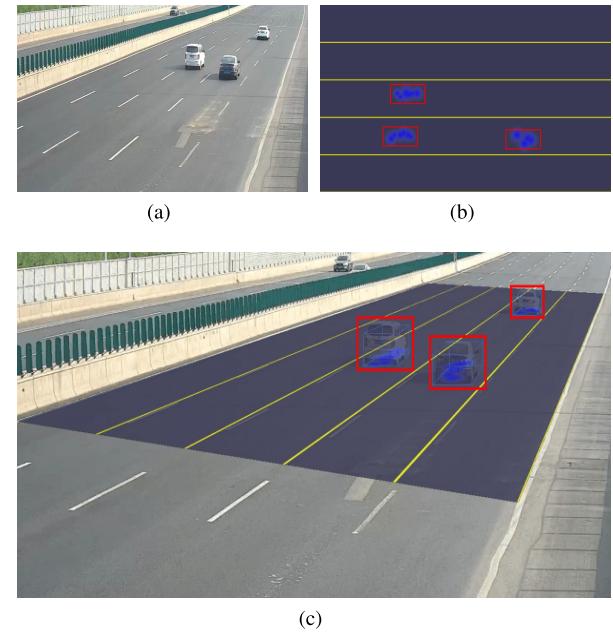
**FIGURE 4.** Moving region detection, (a) camera detection, (b) radar detection, and (c) projection onto the image.

Fig. 4 presents the moving regions obtained by using the proposed method. As shown in Figure 4(b), we can obtain the bird's eye view (BEV) bounding boxes by the K-means method. Areas containing vehicle targets are marked with blue color, and the red boxes in Figure 4(c) are the moving regions. The proposed method is effective when the video and radar data are obtained continuously.

### B. BIDIRECTIONAL REGION DIVISION

To boost the details of different regions in different exposure images, we apply a bidirectional regional partition scheme based on the cumulative histogram analysis of both the original and the reference images. The probability of a pixel having a luminance value  $k$  is expressed as:

$$P_k = \frac{N_k}{N} \quad (3)$$

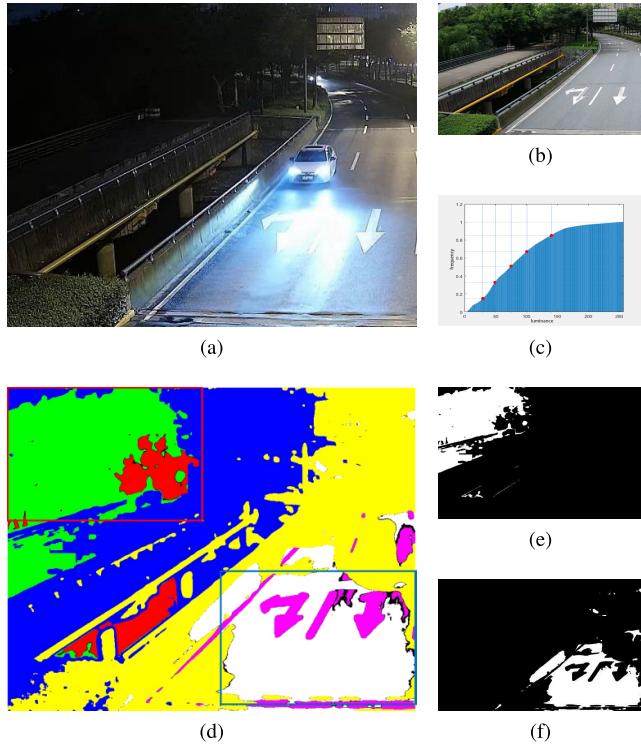
where  $N_k$  is the number of pixels with luminance value  $k$ , and  $N$  denotes the total number of pixels in the image. The accumulated probability up to luminance "m" is expressed as:

$$P_a(m) = \sum_{k=0}^m P_k \quad (4)$$

By analyzing the accumulated probability of the original image, the preliminary region segmentation is based on the following:

$$\begin{cases} \Omega_i = \{p | l(p) \in (l_{i-1}, l_i]\} \\ P_a(l_i) = \frac{i}{R}, \ i = 1, 2, \dots, R \end{cases} \quad (5)$$

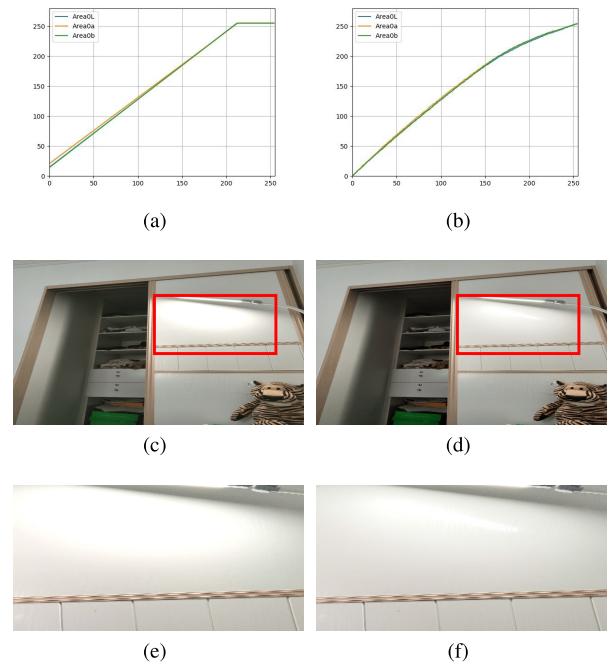
where  $\Omega_i$  denotes the  $i$ -th region, and  $l_i$  is the threshold luminance. This method can divide the image into  $R$  parts,



**FIGURE 5.** Region Division (a) original image, (b) reference image, (c) accumulated histogram of the original image, (d) Regions, (e) local shadow, and (f) local light.

where the first region represents the darkest region, the  $R$ -th region represents the brightest region, and the rest represents the different regions from dark to light. It can be seen that the segmentation depends on the distribution of the image histogram, and therefore it can be adapted to scenes of different brightness distributions.

This work employs the reference image histogram to segment the nighttime image and point out local lighting and shadows. We use the background extraction algorithm for a fixed field-of-view (FOV) camera to obtain a well-exposed image as the reference image [35]. Although our algorithm is not sensitive to background changes, we recommend using the reference image of the same day for night image enhancement to achieve better results. For the darkest and brightest regions, the threshold luminance  $p_1$  and  $p_r$  of these two regions are used to produce the other four regions. These threshold luminance values depend on both the original and the reference images. In this way, local lighting and shadows are effectively extracted. The dark areas in the reference image and the bright regions in the original image are local lighting areas, with Fig. 5 illustrating this effect. The small-scale Gaussian smoothing ( $\sigma = 5$ ) is performed on both the original and the reference images to eliminate the influence of noise. Each local region is represented by a different color in Fig. 5(d), with the threshold luminances shown in Fig. 5(c). Moreover, Fig. 5(e) and Fig. 5(f) demonstrate that our method can accurately point out the local light and



**FIGURE 6.** Result of different iTMO, (a) linear mapping curve, (b) the proposed mapping curve, (c) result of linear method, (d) result of the proposed method, and (e, f) detail magnification.

shadow areas while avoiding confusion with scene textures, such as road marking arrows.

### C. SUBREGIONAL TONE MAPPING

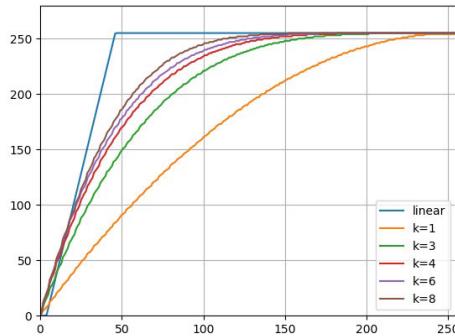
To obtain the image sequence for multi-exposure fusion, we design a mapping scheme based on curve fitting that boosts the details in different regions. Specifically, the statistical characteristics of luminance in the reference image are analyzed, and a non-linear mapping curve is designed to ensure that the region's characteristics in the enhanced image are consistent with the reference image.

The proposed tone mapping function is:

$$\begin{cases} P(t) = \sum_{i=0}^n B_i^n(t)P_i \\ B_i^n(t) = \binom{n}{r} t^i (1-t)^{n-i} \\ \binom{n}{r} = \frac{n!}{i!(n-i)!} \\ P_i = (l_{n,i}, (l_{n,i} - \mu_{n,i})\frac{\sigma_{r,i}}{\sigma_{n,i}} + \mu_{r,i}) \end{cases} \quad (6)$$

where  $P_i$  is the control point of the mapping curve. By designing the control points, we balance the color mapping degree and the texture maintenance. Specifically, this paper divides the brightness range equally into  $M$  portions by  $M+1$  control points. Fig. 6 compares the proposed method ( $k = 6$ ) and the traditional mapping scheme [29].

Fig. 6 (a) presents the three-channel mapping curve obtained by the traditional linear color mapping scheme, and Fig. 6 (b) depicts the mapping curve of the proposed method. Fig. 6 (c) and (e) highlight that linear mapping leads to partial texture loss in the brightness region. As presented in Fig. 6 (d)

**FIGURE 7.** Tone mapping curves with different  $k$ .**TABLE 1.** Metrics of tone-mapped images generated with various  $k$  values.

K	SSIM	E	$\Delta\mu$	$\Delta\sigma$
0	0.6305	5.1344	0.011	0.009
1	0.8716	7.4215	0.254	0.266
3	0.7190	7.1659	0.107	0.092
4	0.7572	6.9747	0.092	0.029
6	0.7366	6.8110	0.074	0.015
8	0.7181	6.6681	0.069	0.007

and the enlarged in Fig. 6 (f), the proposed algorithm avoids the loss of details in high-brightness regions while achieving color migration.

Fig. 7 and Table 1 summarize the statistical characteristics and evaluation metrics with different  $k$  ( $k = 1, 3, 4, 6, 8$ ). The results reveal that our method's statistical characteristics error ( $\Delta\mu$ ) is less than 10%, while the structural similarity (SSIM) and Information Entropy (E) metric are significantly improved. Increasing the number of control points will further decrease the statistical characteristics error. Therefore, we set  $k = 6$  to balance the degree of enhancement and texture maintenance.

A sequence of pseudo-multi-exposure images with boosted details is obtained using the method described above. Fig. 8 presents the mapping curves when the region segmentation level ( $R$ ) is 4, 6, 7, and 8. As the segmentation granularity increases, the image sequence can provide more information for the fusion. Nevertheless, starting from level 7, the overlap occurs on some regional mapping curves for the characteristic relationship between the original image and the reference tends to stabilize in these well-exposed areas. Table 2 highlights that our algorithm's time consumption positively correlates with parameter  $R$ . We find that six pseudo-multi-exposure images are sufficient to produce high perceptual quality images, so we adopt this number in the simulations.

Fig. 9 presents the results of the emulated images with multiple exposures generated from the image “urban”. Fig. 9(a) infers that the proposed method can adaptively generate mapping curves for the images captured in different scenes. Graph (c) is the reference image captured in the daytime.

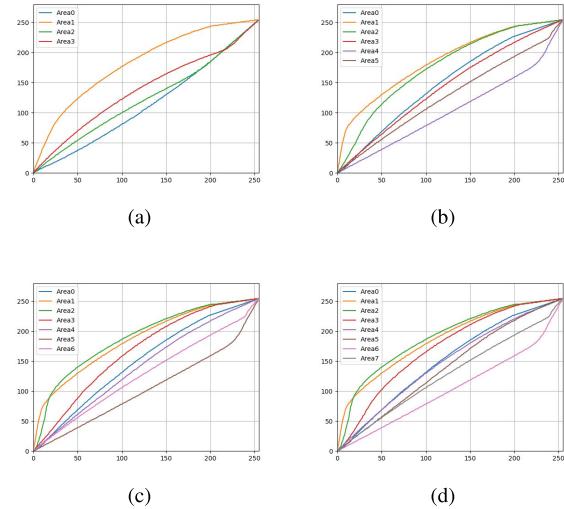
**FIGURE 8.** Tone mapping curves of multi exposures generated by different  $R$  ( $K = 5$ ), (a)  $R = 4$ , (b)  $R = 6$ , (c)  $R = 7$ , and (d)  $R = 8$ .**TABLE 2.** Time comparison under different  $R$  ( $K = 4$ ).

Image	R	Time(s)(400 × 400)
		Urban
	4	0.21
	5	0.26
	6	0.34
	7	0.38

Besides, Fig. 9(d) highlights that details of different regions are boosted in different exposure images.

#### D. MOVING REGION ENHANCEMENT

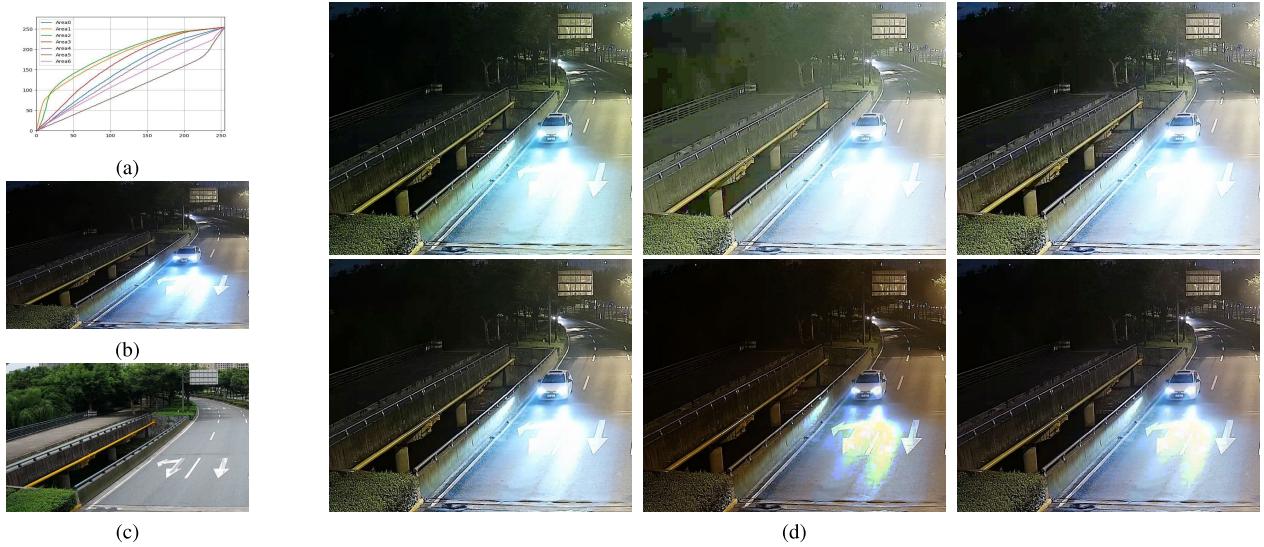
The above pseudo-exposure image generation method is unsuitable for areas with moving objects. Key information, such as pedestrians and vehicles, may be faded due to the tone mapping of the reference image. In this paper, we perform a PDE-based algorithm to realize contrast enhancement and gradient-domain reconstruction synchronously, represented as follows:

$$\begin{cases} \frac{\partial I}{\partial t} = \frac{D_{max}-I}{D_{max}-D_{min}} A_{\Omega} - A(I) \\ A(v) = Area[I \geq v] \\ \frac{\partial \hat{I}}{\partial t} = \Delta \hat{I} - div(I) \end{cases} \quad (7)$$

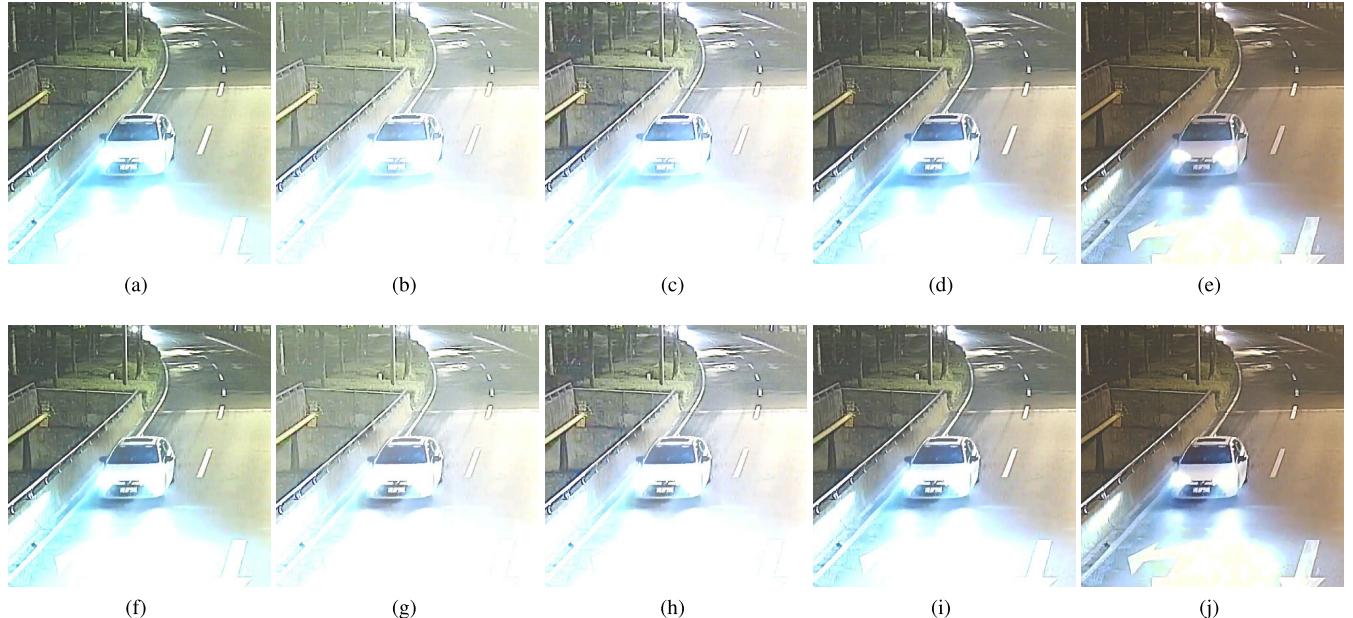
where  $(D_{max}, D_{min})$  denote the maximum and minimum luminance values of the pixels in the area  $\Omega$ .  $A(v)$  represents the area (or the number of pixels in the discrete case), and  $A_{\Omega}$  is the total number of pixels in area  $\Omega$ . The steady-state solution is given by:

$$I_{k+1} = I_k + \Delta t((D_{max} - D_{min})H(I_k) - I_k) \quad (8)$$

where  $H(I_k)$  is the accumulated histogram of the  $k$ -th iteration of the input image.  $\Delta t$  denotes the time step. Besides, it should be noted that the algorithm degenerates into traditional histogram equalization for  $\Delta t = 1$ .



**FIGURE 9.** Examples of multi-exposure images generated from the image “urban” with  $R = 6, K = 6$ .



**FIGURE 10.** Enhancement of moving region, (a-e) moving region directly generated by tone mapping, (f-j) after enhanced by PDE-HM.

According to the continuity of the boundary with the stationary region, the Dirichlet boundary conditions can be added:

$$\hat{I}|\partial\Omega = I_{in}|\partial\Omega \quad (9)$$

Finally, the proposed fast PDE-based histogram modification method is summarized in Algorithm 1.

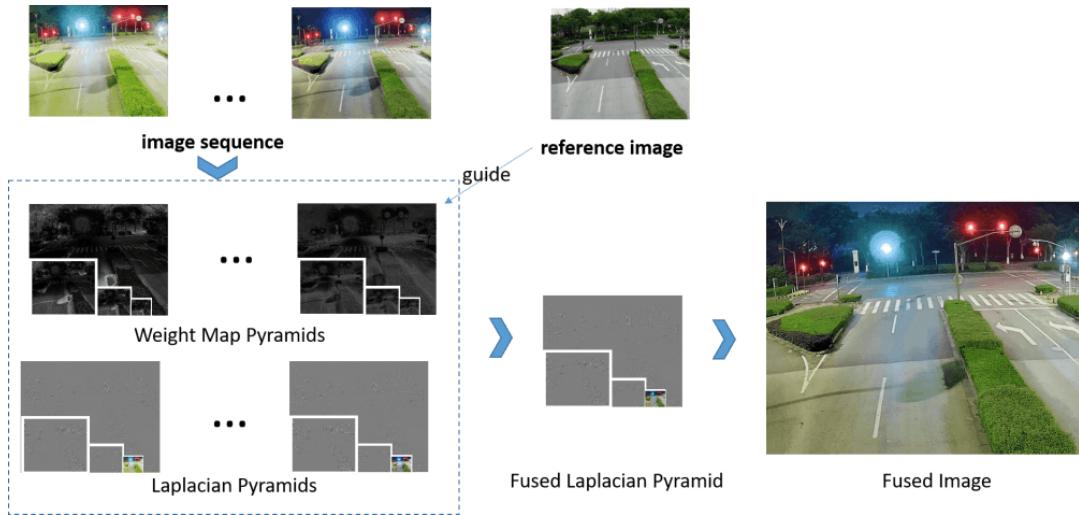
Fig. 10 compares the results before and after using the PDE-HM method, revealing that the car in Fig. 10 (a), generated directly by tone mapping, is faded. Fig. 10 (b) highlights that the proposed PDE-HM method can boost the details of

traffic targets while avoiding the unnatural seams between the moving and stationary regions.

#### E. MULTI-SCALE FUSION

We use an improved weighted Laplace pyramid fusion algorithm to fuse the obtained image sequence. An overview of the Multi-scale fusion technique is provided in Fig. 11. To construct the fused Laplacian pyramid, we fuse multiple images from each exposure using the following equation:

$$\begin{cases} \hat{W}_{i,k}(p) = \frac{W_{i,k}(p)+\epsilon}{\sum_{t=1}^R (W_{i,t}(p))} \\ LF_i(p) = \sum_{k=1}^R \hat{W}_{i,k}(p)L_{i,k}(p) \end{cases} \quad (10)$$



**FIGURE 11.** The flow chart of the proposed Pyramids fusion with guidance information.

#### Algorithm 1 PDE Based Histogram Modification

**Input:** RGB images  $I_{in}$ , the h-th image of multi-exposures ( $I_h$ )  
**Output:** Enhanced image ( $\hat{I}$ )

---

```

1: Initialize Set  $I_0 = I_{in}$ ,  $\hat{I}_0 = I_{in}$ 
2: While  $\epsilon_I > \eta$ ,  $\epsilon_{\Delta I} > \tau$ ,  $k < K_{max}$  do:
3:   Calculate accumulated histogram of k-th image  $H(I_k)$ 
4:    $I_{(k+1)} \leftarrow I_k + \Delta t((D_{max} - D_{min})H(I_k) - I_k)$ 
5:   Calculate  $\Delta(\hat{I}_k)$  with dirichlet boundary Equation (9)
6:    $\hat{I}_{k+1} \leftarrow \hat{I}_k + \Delta t(\Delta(\hat{I}_k) - \text{div}(I_k))$ 
7:    $\epsilon_I \leftarrow \|I_{k+1} - I_k\|_2^2$ 
8:    $\epsilon_{\Delta I} \leftarrow \|\hat{I}_{k+1} - \hat{I}_k\|_2^2$ 
9: End
10: The enhanced image  $\hat{I} = \hat{I}_k$ 
```

---

where  $LF_i(p)$  is the weighted fusion result of the  $i$ -th layer of the Laplacian pyramid.  $L_{i,k}(p)$  is the  $k$ th sub-Laplace pyramid that can be obtained layer by layer from the  $k$ th image using a sampling algorithm.  $\hat{W}_{i,k}(p)$  is the normalized weight map of the  $k$ th image, and  $W_{i,k}(p)$  denotes the weight map obtained as follows:

$$W_{p,k} = C_{p,k}^{\omega_c} \times S_{p,k}^{\omega_s} \times E_{p,k}^{\omega_e} \quad (11)$$

where  $(\omega_c, \omega_s, \omega_e)$  are adjustment parameters, and  $(C_{p,k}, S_{p,k}, E_{p,k})$  are the contrast weight components, saturation weight component, and exposure weight components, respectively, which are defined as:

$$\begin{cases} C_{p,k} = L(I_{gray}) \\ S_{p,k} = \exp\left(\frac{-(\hat{s}_{p,k} - \hat{s}_{p,g})^2}{2\sigma_s^2}\right) \\ E_{p,k} = \prod_{c=r,g,b} \exp\left(\frac{-(l_{p,k}^c - l_g^c)^2}{2\sigma_E^2}\right) \end{cases} \quad (12)$$

where  $L()$  is a Laplacian filter,  $I_{gray}$  is the grayscale of the input image.  $l_{p,k}$  and  $l_g$  denote the luminance.  $\hat{s}_{p,k}$  and  $\hat{s}_{p,g}$  are the saturation weights of the  $k$ -th image of the multi-exposures and the reference image at this pixel,

respectively. Besides, it should be noted that  $l_g$  is the guidance information generated by the reference image that are generated by applying a large-scale Gaussian to the reference image, or the image structure information decomposed by:

$$l_g = \arg \min\left(\frac{1}{2}\|l - i_r\|_2^2 + \alpha\|\nabla l\|_1\right) \quad (13)$$

where  $i_r$  is the reference image brightness, and  $\|\cdot\|_1$  is the 1 norm.

Finally, the pyramid  $LF_i(p)$  is collapsed to obtain a fusion image  $F_0(p)$ :

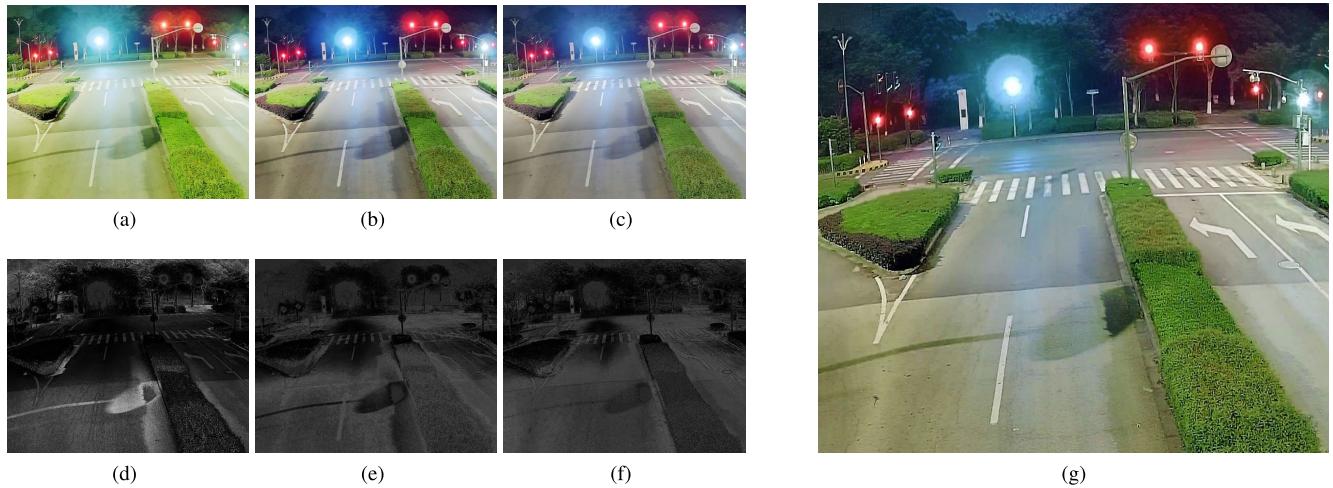
$$F_i(p) = \begin{cases} LF_i(p) + \text{PyrUp}(LF_{i+1}(p)) & 0 \leq i < N - 1 \\ LF_i(p) & i = N - 1 \end{cases} \quad (14)$$

where  $\text{PyrUp}()$  denotes an up-sampling function and  $F_0(p)$  is the final enhanced image.

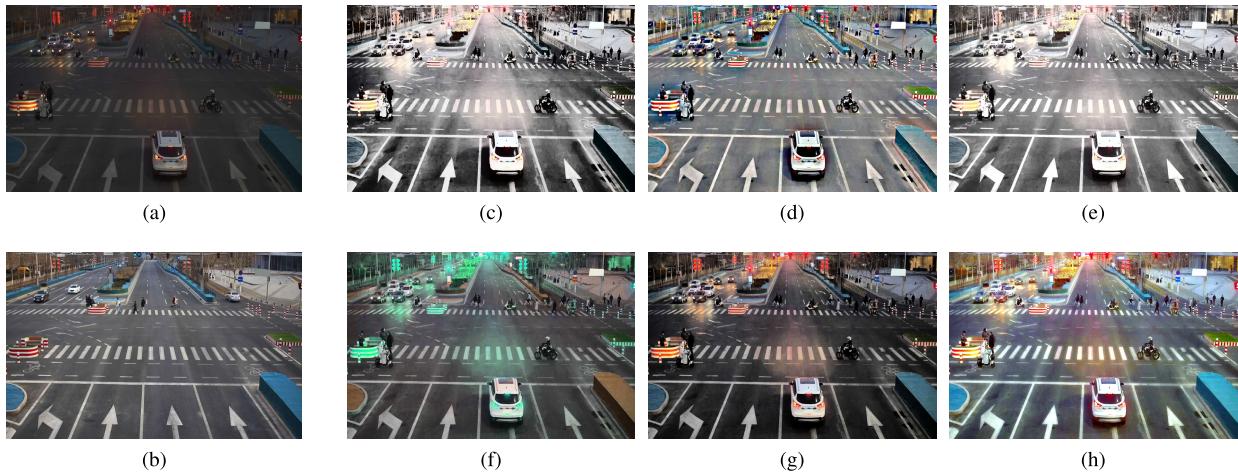
Fig. 12 depicts a layer's weighted fusion result, revealing that our algorithm's weight map has continuous values. The visual effect of the fused image is much better than any of the original multi-exposure images, as the details are retained comprehensively. Compared with traditional MEF, our method shows great advantages in the naturalness of exposure and color due to the addition of guidance information. The overall workflow of the proposed method is presented in Algorithm 2.

#### IV. EXPERIMENTAL RESULTS

This section evaluates our night color image enhancement method and car detection in night traffic scenes. We compare the proposed algorithm against the dark channel prior (DCP) based luminance stretching (G. Sahu, 2019 [31]), Retinex-based method (A. Mgg, 2020 [32]), MEF-based methods (Xu, K. et. al. 2022 [14]), and Deep Learning based methods (GAN-Fuse by Z. Yang, Et. 2021 [15] and DeepFuse by Prabhakar.



**FIGURE 12.** Weighted fusion of images, (a) – (c) Multi exposure images, (d) – (f) Weight maps, (g) Fused image.



**FIGURE 13.** Results on the “Crossroads”, (a) source image, (b) reference image, (c) DCP, (d) MSRCR, (e) MEF (paper [14]), (f) DeepFuse (paper [17]), (g) GANFuse (paper [15]), (h) Proposed.

Et. 2017 [17]). All methods are quantitatively evaluated through several metrics.

#### A. DATA SETS AND EXPERIMENT SETTING

We select test images from the Rope3D dataset, which is remarkably known for its spacious range of features toward C-ITS scenes. The dataset images are obtained from the Intelligent Roadside Surveillance System (IRSS), and the data set comprises three parts. The first part contains the desensitized raw images captured by traffic surveillance cameras in diverse environments and scenes (day/night, sunny/rainy, and urban/suburban areas), and the second part contains the annotated files, timestamp files, and calibration files. Finally, the last part contains the corresponding radar detection data of the first part [34].

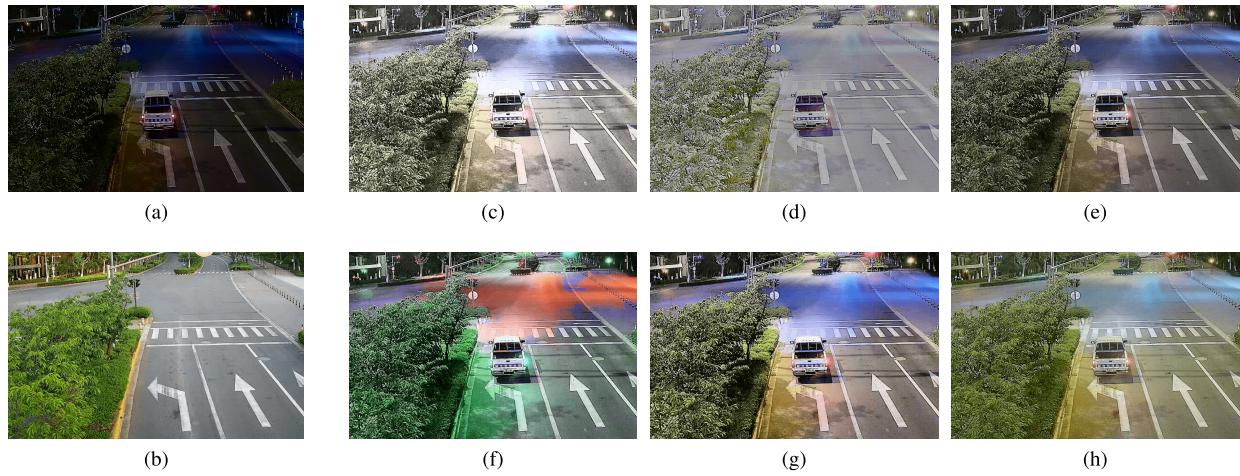
The computing platform for the experiment is the Intelligent Roadside Surveillance System. The core processing module is a compact customized industrial computer with a 64-bit Ubuntu16.04 operating system, utilizing an Inter (R)

Core i7-7800x CPU 3.50GHz, 16GB memory, and a GeForce GTX 2080Ti. The software is programmed in C++.

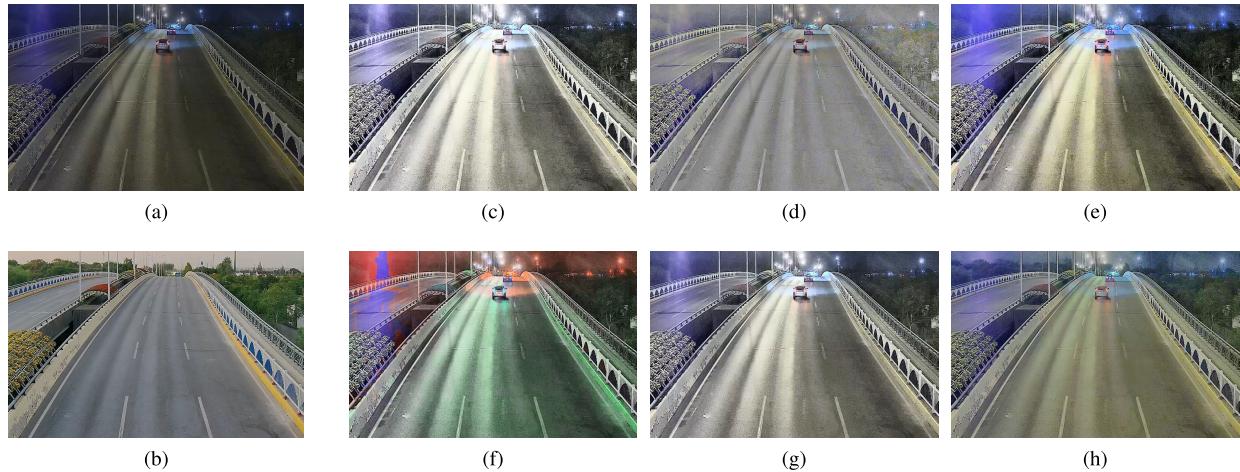
#### B. SUBJECTIVE ASSESSMENT

To verify the proposed method’s effectiveness, we show visual examples of three typical traffic scenes: (a) the urban roads (Urban Rd), which suffer light spots and shadows caused by street lamps and trees, (b) pedestrians crossing at night (Crossroads) is a typical scene with traffic lights and signs, and (c) the poorly lit bridge (Bridge), where the natural lighting is insufficient, with the car headlights being the main light source. The corresponding results are illustrated in Fig. 13 –15. The reference images, i.e., graph (b) of each image group, are obtained by processing the background extraction algorithm during the daytime [35].

Fig. 13 is an example of images captured in the Crossroads scene. It can be seen that DCP can enhance the contrast, but details are lost in the area close to street lamps. The



**FIGURE 14.** Results on the “Urban Rd”, (a) source image, (b) reference image, (c) DCP, (d) MSRCR, (e) MEF (paper [14]), (f) DeepFuse (paper [17]), (g) GANFuse (paper [15]), (h) Proposed.



**FIGURE 15.** Results on the “Bridge”, (a) source image, (b) reference image, (c) DCP, (d) MSRCR, (e) MEF (paper [14]), (f) DeepFuse (paper [17]), (g) GANFuse (paper [15]), (h) Proposed.

MEF-based method balanced the bright and dark areas by adaptively boosting different details in the light and dark areas, but the vehicle details in the upper left corner deteriorated severely. The results of MSRCR and the proposed method are visually pleasing. Compared with the other five approaches, our method presents the best performance.

Fig. 14 reveals that DCP fails in the enhancement, for it equates detail enhancement with visual quality improvement, which is not always true, especially in the case of over-enhancement. The result of deepfuse has color deviation. MSRCR can deal with the light spots and shadows, while the dynamic range of the result image is so small that the visibility is disappointing. Moreover, although GANFuse’s method can successfully enhance the image’s details, the road ahead of the car showed an unnatural blue color. As shown in Fig. 14(h), our method affords the best performance in the naturalness of exposure and color due to adding guidance information.

Fig. 15 visually compares our method with the competitor algorithms on the image sequence “Bridge”. The results infer

that DCP performs better than in the other two scenes, while the area, including the car, is still over-enhanced. The noise in the result of MSRCR is obvious. Compared with the results of competitor methods, the color in our method is more natural.

### C. OBJECTIVE ASSESSMENT

As the subjective evaluation is easily affected by personal factors, we assess the performance of the proposed method objectively using four quality metrics: Image Enhancement Measure Evaluation (EME) [37], CIEDE2000 [38], HDR-VDP-2 [39], and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [40], that evaluate the proposed method from the aspects of enhanced image quality, similarity with the structure of the original image, and color restoration degree. The results are reported in Table 3.

EME, given by Equation (15), approximates the average contrast in an image by dividing it into blocks. The larger the EME value, the better the image quality after enhancement.

**TABLE 3.** Objective metrics' scores of enhanced images generated by different method.

Metrics	Image	DCP	MSRCR	Paper [14]	Paper [17]	Paper [15]	Proposed
BRISQUE	Crossroads	29.07	26.64	22.18	25.65	26.91	<b>20.98</b>
	Urban Rd	29.73	20.77	19.06	19.82	20.55	<b>18.63</b>
	Bridge	22.20	<b>12.19</b>	17.52	16.41	16.95	<b>12.39</b>
EME	Crossroads	46.11	44.20	48.77	54.20	<b>58.11</b>	49.01
	Urban Rd	58.42	54.35	64.81	69.26	<b>70.83</b>	64.28
	Bridge	50.62	42.14	46.95	<b>55.03</b>	45.68	48.12
HDR-VDP-2	Crossroads	38.10	39.89	38.12	37.24	36.66	<b>40.21</b>
	Urban Rd	41.30	43.63	41.53	41.54	41.03	<b>44.20</b>
	Bridge	42.61	<b>49.33</b>	46.99	45.17	42.90	<b>47.97</b>
CIE	Crossroads	24.12	19.88	30.45	27.76	29.21	<b>17.24</b>
	Urban Rd	26.58	20.90	31.61	27.35	29.73	<b>16.19</b>
	Bridge	22.12	19.40	21.30	20.76	18.33	<b>12.47</b>

**Algorithm 2** The Overall Workflow of the Proposed Method

**Input:** Night color image  $I$ , the reference image  $R$ , the radar detections  $Obj_{radar}$   
**Output:** Enhanced image  $I_{en}$

- 1: **Initialize:** set  $I_{en} = I$ , set the region segmentation level  $rank = 6$ , set the pyramid decomposition level  $level = 3$
- 2: **Calculate moving regions:**
- 3: **for** each obj  $\in Obj_{radar}$
- 4:     Calculate ROI ( $Mask_{ROI}$ ) in image by equation (2)
- 5: **end for**
- 6: **Calculate multi-exposures:**
- 7: **for** each  $i \in [0, rank]$  **do:**
- 8:     Calculate mapping region ( $Mask_i$ ) by equation (5)
- 9:     Calculate tone mapping curve in mapping region by equation (6)
- 10:    Calculate  $i$ th exposure image by tone mapping curve
- 11: **Enhance moving regions:**
- 12:   **for**  $p \in Mask_{ROI}$  **do:**
- 13:     Calculate enhanced pix  $\hat{p}$  by **Algorithm 1**
- 14:   **end for**
- 15: **end for**
- 16: **Gaussian pyramid decomposiong:**
- 17: **for** each image  $i \in [0, rank]$  **do:**
- 18:     Calculate gaussian pyramid decomposiong:
- 19: **end for**
- 20: **Guided laplace pyramid fusion:**
- 21: **for**  $i \in [0, level]$  **do:**
- 22:     **for**  $j \in [0, rank]$  sequences **do:**
- 23:         Calculate fusion weight by equation (11)
- 24:     **end for**
- 25:     Calculate fused level image by equation (10)
- 26: **end for**
- 27: **for**  $i \in \{level - 1 \dots 0\}$  **do:**
- 28:      $F_i = \begin{cases} LF_i + PyrUp(LF_{i+1}) & 0 \leq i < level - 1 \\ LF_i & i = level - 1 \end{cases}$
- 29: **end for**
- 30: The enhanced image  $I_{en} = F_0$

We set  $M = 20$ ,  $N = 20$  in this paper.

$$EME = \frac{1}{M \times N} \sum_{l=1}^M \sum_{k=1}^N 20 \log \frac{l_{\max,k,l}^w}{l_{\min,k,l}^w} \quad (15)$$

CIEDE2000 (CIE) describes the color difference between the enhanced image and the natural scene during the day. Due to the existence of moving objects, this paper only counts the color deviation index of the stationary region, calculated by:

$$CIE = \frac{1}{M \times N} \sqrt{\sum_{c=l,a,b} (c_1 - c_2)^2} \quad (16)$$

The BRISQUE first extracts the mean subtracted contrast normalized coefficients, calculates the feature vector, and uses a support vector machine to perform regression. The lower the image quality score, the better the image quality.

DR-VDP-2 evaluates the similarity between images based on the human visual system (HVS). A higher HDR-VDP-2 score denotes a smaller visual difference between the pair. Unfortunately, most traffic surveillance cameras do not provide HDR images, which are used to calculate the HDR-VDP-2 scores. We use the reference image instead of the nighttime HDR image to overcome this issue.

Table 3 highlights that the proposed method has an appealing performance, although not the best, as the CIE scores it obtains are generally minimal. This is because the proposed method utilizes extra color information from the reference image. The EME scores show that our algorithm increases the image texture information to a certain extent, but it is not optimal. This is because we focus on the naturalness of the image than stretching contrast. Our method affords the best performance in BRISQUE and HDR-VDP-2 scores, describing the naturalness of exposure and color.

**D. EFFICIENCY COMPARISON AND ANALYSIS**

We analyze the efficiency of the competitor methods by comparing the average time performing multiple image enhancement tasks. The state-of-the-art (SOTA) algorithm of both MEF and non-MEF are cited for comparison.

Table 4 reports the average processing time of each method, highlighting that the histogram-based method (HE), DCP and Adapter Gamma algorithms (AG) are more processing efficient because they only use simple mappings. The MSRCR algorithm imposes a greater processing burden due to the large-scale Gaussian operation. Although the deep-learning-based methods are efficient for image enhancement,



**FIGURE 16.** Results of traffic target detection, (a) Precision-Recall curve of car detection by different methods, (b) average precisions of target detection, (c) car detection on original image, (d) car detection on enhanced image.

**TABLE 4.** Time comparison of different method.

Method	Aver Time (s)	
	(600 × 600)	(1920 × 1080)
HE	0.07	0.19
DCP	0.12	0.45
MSRCR	1.2	3.1
AG [33]	0.51	0.96
DeepFuse [17]	0.19	0.36
GANFuse [15]	0.21	0.51
MEF [14]	0.66	1.21
MEF [19]	0.78	1.42
Proposed	0.62	1.31

the training work is time-consuming and heavily dependent on laborious attempts and accurate annotation. Our method requires less time than the other ghosting-free MEF algorithms because we use radar information to generate moving regions, thus avoiding many pixel operations.

#### E. APPLICATION TO TRAFFIC TARGET DETECTION

We tested our enhancement method in the nighttime traffic scene's multi-object detection and recognition task.

Fig. 16(c) infers that the original perception image has low visibility, and the deep learning module has not detected all vehicles in the figure. Opposing, the image enhanced by the proposed algorithm, shown as Fig. 16(d), is then subjected to the deep learning target detection module, and therefore, the recognition accuracy is increased significantly.

As shown in Fig. 16(a) (b), when using our method, the traffic target detection system shows a lower miss rate and higher average precision (AP), which is expected for it transforms the color characteristics of the nighttime images to that of the well-exposed images. As an efficient image preprocessing operator, our method can improve the performance of object detection tasks in various luminance conditions without additional training.

#### V. CONCLUSION

To facilitate multi-object detection and recognition tasks on the Intelligent Roadside Surveillance System, we propose an efficient flowchart for night image enhancement by employing multi-source data fusion.

We present a complete set of multi-sensor fusion schemes suitable for the intelligent roadside system. The two key fusions are the decision-level fusion of camera and radar and the pixel-level fusion of day and night images.

By splitting the nighttime image into moving and stationary regions, specialized iTMO strategies are used in different regions to boost the details of the image. Furthermore, pseudo-multi-exposures are generated by an improved region-based tone mapping scheme. In addition, we propose a novel multi-exposure fusion method by considering the well-exposed image as guidance information.

The experimental results reveal that the proposed method can also obtain promising performance to enhance LDR nighttime images and greatly facilitate object (car) detection in nighttime traffic scenes.

The main drawback of the proposed method is that it does not consider the noise suppression problem. This problem can be solved by adding a noise-removal step, such as BM3D. A possibly interesting work is to fuse the reference image, the nighttime image, and a virtual image with brightened moving objects. This will be studied in our future research.

#### REFERENCES

- [1] P. D. Macheret, R. R. Savchuk, and G. I. Shkuratov, "Intelligent transport systems: Analysis of the current state and prospects of development," in *Proc. Int. Conf. Quality Manage., Transp. Inf. Secur., Inf. Technol.*, Sep. 2021, pp. 234–237.
- [2] J. Mao, D. Hong, X. Wang, C. Hsu, and A. Shanthini, "CVRSS-CHD: Computer vision-related roadside surveillance system using compound hierarchical-deep models," *IET Intell. Transp. Syst.*, vol. 14, no. 11, pp. 1353–1362, Nov. 2020.
- [3] T. Soumya and S. M. Thampi, "Nighttime visual refinement techniques for surveillance video: A review," *Multimedia Tools Appl.*, vol. 78, pp. 32137–32158, Jul. 2019.

- [4] G. Sahu, A. Seal, D. Bhattacharjee, M. Nasipuri, P. Brida, and O. Krejcar, "Trends and prospects of techniques for haze removal from degraded images: A survey," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 4, pp. 762–782, Aug. 2022.
- [5] G. Sahu, A. Seal, O. Krejcar, and A. Yazidi, "Single image dehazing using a new color channel," *J. Vis. Commun. Image Represent.*, vol. 74, Jan. 2021, Art. no. 103008.
- [6] G. Sahu, A. Seal, A. Yazidi, and O. Krejcar, "A dual-channel dehaze-net for single image dehazing in visual Internet of Things using PYNNQ-Z2 board," *IEEE Trans. Autom. Sci. Eng.*, early access, Nov. 1, 2022, doi: 10.1109/TASE.2022.3217801.
- [7] J. Bai, S. Li, L. Huang, and H. Chen, "Robust detection and tracking method for moving object based on radar and camera data fusion," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10761–10774, May 2021.
- [8] P. Liu, G. Yu, Z. Wang, B. Zhou, and P. Chen, "Object classification based on enhanced evidence theory: Radar-vision fusion approach for roadside application," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [9] M. Gao, J. Wang, Y. Chen, C. Du, C. Chen, and Y. Zeng, "An improved multi-exposure image fusion method for intelligent transportation system," *Electronics*, vol. 10, no. 4, p. 383, Feb. 2021.
- [10] M. Bouain, D. Berdag, N. Fakhfakh, and R. B. Atallah, "Multi-sensor fusion for obstacle detection and recognition: A belief-based approach," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Jul. 2018, pp. 1217–1224.
- [11] Y. Li, M. Liu, and K. Han, "Overview of multi-exposure image fusion," in *Proc. Int. Conf. Electron. Commun., Internet Things Big Data (ICEIB)*, Dec. 2021, pp. 196–198.
- [12] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 64, pp. 71–91, Dec. 2020.
- [13] F. Xu, J. Liu, Y. Song, H. Sun, and X. Wang, "Multi-exposure image fusion techniques: A comprehensive review," *Remote Sens.*, vol. 14, no. 3, p. 771, Feb. 2022.
- [14] K. Xu, Q. Wang, H. Xiao, and K. Liu, "Multi-exposure image fusion algorithm based on improved weight function," *Front. Neurorobot.*, vol. 16, Mar. 2022, Art. no. 846580.
- [15] Z. Yang, Y. Chen, Z. Le, and Y. Ma, "GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 6133–6145, Jun. 2021.
- [16] O. Ulucan, D. Karakaya, and M. Turkan, "Multi-exposure image fusion based on linear embeddings and watershed masking," *Signal Process.*, vol. 178, Jan. 2021, Art. no. 107791.
- [17] K. R. Prabhakar, V. S. Srikanth, and R. V. Babu, "DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4724–4732.
- [18] S. Sen Gupta, S. Hossain, and K. Kim, "HDR-like image from pseudo-exposure image fusion: A genetic algorithm approach," *IEEE Trans. Consum. Electron.*, vol. 67, no. 2, pp. 119–128, May 2021.
- [19] Y. Kinoshita, T. Yoshida, S. Shiota, and H. Kiya, "Pseudo multi-exposure fusion using a single image," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2017, pp. 263–269.
- [20] T. Wang, C. Chiu, W. Wu, J. Wang, C. Lin, C. Chiu, and J. Liou, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 470–484, Apr. 2015.
- [21] X. Wang, R. Hu, and X. Xu, "Single low-light image brightening using learning-based intensity mapping," *Neurocomputing*, vol. 508, pp. 315–323, Oct. 2022.
- [22] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7448–7458, Dec. 2020.
- [23] D. Liu, Y. Cui, L. Yan, C. Mousas, and Y. Chen, "DenserNet: Weakly supervised visual localization using multi-scale feature aggregation," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 1–9.
- [24] M. Zhang, T. Liu, Y. Piao, S. Yao, and H. Lu, "Auto-MSFNet: Search multi-scale fusion network for salient object detection," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 667–676.
- [25] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.
- [26] K. R. Prabhakar and R. V. Babu, "Ghosting-free multi-exposure image fusion in gradient domain," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1766–1770.
- [27] M. Aeberhard and T. Bertram, "Object classification in a high-level sensor data fusion architecture for advanced driver assistance systems," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 416–422.
- [28] F. Vasconcelos, J. P. Barreto, and U. Nunes, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2097–2107, Nov. 2012.
- [29] E. Reinhard, M. Adhikmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 34–41, Jul. /Aug. 2001.
- [30] M. F. Al-Sammaraie, "Contrast enhancement of roads images with foggy scenes based on histogram equalization," in *Proc. 10th Int. Conf. Comput. Sci. Educ. (ICCSCE)*, Jul. 2015, pp. 95–101.
- [31] G. Sahu and A. Seal, "Image dehazing based on luminance stretching," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Dec. 2019, pp. 388–393.
- [32] M. G. Girija, K. T. Shanavaz, and G. S. Ajith, "Image dehazing using MSRCA algorithm and morphology based algorithm: A concise review," *Mater. Today, Proc.*, vol. 24, pp. 1890–1897, Jan. 2020.
- [33] K. Yang, H. Li, H. Kuang, C. Li, and Y. Li, "An adaptive method for image dynamic range adjustment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 640–652, Mar. 2019.
- [34] X. Ye, M. Shu, H. Li, Y. Shi, Y. Li, G. Wang, X. Tan, and E. Ding, "Rope3D: The roadside perception dataset for autonomous driving and monocular 3D object detection task," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21309–21318.
- [35] A. Tavakkoli, M. Niculescu, and G. Bebis, "Automatic robust background modeling using multivariate non-parametric kernel density estimation for visual surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2005, pp. 246–252.
- [36] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.
- [37] S. S. Agaian, K. P. Lenz, and A. M. Grigoryan, "A new measure of image enhancement," in *Proc. IASTED Int. Conf. Signal Process. Commun.*, 2000, pp. 19–22.
- [38] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Res. Appl.*, vol. 30, no. 1, pp. 21–30, Feb. 2005.
- [39] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, Jul. 2011.
- [40] A. Mittal and A. K. Moorthy, "Blind/referenceless image spatial quality evaluator," in *Proc. Signals, Syst. Comput. (ASILOMAR)*, 2012, pp. 723–727.



**CHEN YAQING** (Member, IEEE) received the M.S. degree in control science and engineering from the Beijing University of Aeronautics and Astronautics, Peking, in 2012. He is currently pursuing the Ph.D. degree in mechanical engineering with the Nanjing University of Aeronautics and Astronautics, Jiangsu, China. His research interests include image processing, machine vision, and cooperative intelligent transport systems.



**WANG HUAMING** received the M.S. and Ph.D. degrees in mechanical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2001 and 2004, respectively. Since 2004, he has been with the Department of Mechatronics Engineering, Nanjing University of Aeronautics and Astronautics, where he is currently a Professor. In 2013, he was a Visiting Scholar with the Department of Mechanical Engineering, The University of Melbourne. His research interests include hopping robots and dielectric elastomer actuators and soft. He is mainly engaged in the research of image processing and machine vision, electromechanical system design and control, and mobile robot.