

Low-Illumination Image Enhancement for Night-Time UAV Pedestrian Detection

Weijiang Wang , Yeping Peng , Member, IEEE, Guangzhong Cao , Senior Member, IEEE, Xiaoqin Guo, and Ngaiming Kwok 

Abstract—To accomplish reliable pedestrian detection using unmanned aerial vehicles (UAVs) under night-time conditions, an image enhancement method is developed in this article to improve the low-illumination image quality. First, the image brightness is mapped to a desirable level by a hyperbolic tangent curve. Second, the block-matching and 3-D filtering methods are developed for an unsharp filter in YCbCr color space for image denoising and sharpening. Finally, pedestrian detection is performed using a convolutional neural network model to complete the surveillance task. Experimental results show that the Minkowski distance measurement index of enhanced images is increased to 0.975, and the detection accuracies, in F-measure and confidence coefficient, reach 0.907 and 0.840, respectively, which are the highest as compared with other image enhancement methods. This developed method has potential values for night-time UAV visual monitoring in smart city applications.

Index Terms—Block-matching and 3-D filtering, hyperbolic tangent curve (HTC), low-illumination image enhancement, night-time detection, unmanned aerial vehicle (UAV).

I. INTRODUCTION

VISUAL inspection technology is widely applied in industries and many other applications. For smart city management, unmanned aerial vehicles (UAVs) can be integrated with computer vision techniques to automatically detect vehicles and pedestrians [1], [2]. To obtain reliable detections in unfavorable conditions, such as night-time or overcast weather,

Manuscript received March 26, 2020; revised June 2, 2020 and August 27, 2020; accepted September 18, 2020. Date of publication September 23, 2020; date of current version May 3, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 51905351 and Grant U1813212, in part by the Natural Science Foundation of Guangdong Province, China, under Grant 2018A030310522, and in part by the Science and Technology Planning Project of Shenzhen Municipality, China, under Grant JCYJ2019080813413430. Paper no. TII-20-1545. (Corresponding author: Yeping Peng.)

Weijiang Wang, Yeping Peng, Guangzhong Cao, and Xiaoqin Guo are with the Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: 1810294026@email.szu.edu.cn; pyp8020@163.com; gzcao@szu.edu.cn; guoxq@szu.edu.cn).

Ngaiming Kwok is with the School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW 2052, Australia (e-mail: drnmkwok@hotmail.com).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2020.3026036

low-illumination image enhancement is a crucial requirement that has to be satisfied [3], [4].

Image enhancement is one of the main concerns if an intended vision-based task can be accomplished or not. These include video surveillance, robotics, medical science, and machine condition monitoring [5]–[9]. In order to improve the quality of underwater images, Lu *et al.* [10] proposed an image reconstruction model built by a deep convolutional neural network. An image enhancement method that adjusts image luminance and chrominance to deal with the low-illumination problem was also reported [11]. The medical images were transformed into two subbands of low and high frequencies for image enhancement [12]. Besides, image enhancement is important for intelligent transportation applications and visual-aided driving [13].

The above image enhancement methods are mostly applied in static or specific scenes, and the processing results depend critically on illumination conditions. However, outdoor UAV surveillance is influenced by flight operations, complex environment backgrounds, and illumination conditions, making it difficult to detect the target objects at night. To this end, researchers began to investigate night-time image enhancement methods, as shown in Table I. These algorithms cannot be directly adopted in UAV image processing because of excessive enhancement [14], artefact noise [15]–[18], excessive smooth [19], and complicated computation [20]–[22].

Deep learning algorithms have been widely applied in low-light image enhancement [23]–[26]. However, deep learning methods are also unsuitable for the real-time processing of the UAV images. It can be seen in Table I that paired training with public dataset images under different illuminations is necessary for the LLnet, the KinD, and the RDGAN network models. In practice, it is difficult to control a flying UAV to capture two images with the same scene in day and night. Although the EnlightenGAN can solve the problem of paired training, its complicated computation and long processing time cannot meet the real-time requirement in UAV surveillance systems.

This article aims at developing an efficient and robust image enhancement method for night-time UAV image processing. The hyperbolic tangent curve (HTC), block-matching and 3-D filtering (BM3D) algorithms are combined to improve the image quality. The HTC is used to increase the image brightness through weighted brightness stretching. In this stage, the weighting parameter is determined according to the maximum image entropy. Afterward, an unsharp filter is utilized in the BM3D algorithm and in the YCbCr color space, where image denoising

TABLE I
DISADVANTAGES OF NIGHT-TIME IMAGE ENHANCEMENT ALGORITHMS ADOPTED TO PROCESS UAV IMAGES

No.	Algorithm	Published year	Working principle	Disadvantages for UAV image processing
1	Adaptive attenuation quantification Retinex (AAQR) [14]	2018	The Retinex theory was adopted to adaptively attenuation quantify the details of night-time images	May cause excessive enhancement
2	Adaptive Gamma correction with colour preserving framework (AGCCPF) [15]	2016	The Gamma correction and the weighted cumulative probability distribution function were combined to enhance the images	Produce artefact noise
3	Multiderivation fusion (MF) [16]	2016	Multi-order derivatives of the image light map were extracted to manipulate the image brightness	Produce artefact noise
4	Multi-scale Retinex with colour restoration (MSRCR) [17]	2015	The light reflection and illumination was separated by multiple Gaussian filters for the image restoration	Produce artefact noise
5	Physical lighting model (PLM) [18]	2019	A physical lighting model was established to describe the degradation of low illumination images	Produce artefact noise
6	Low-light image enhancement (LIME) [19]	2016	A defogging conversion map model was proposed to optimize the image lighting map	May cause excessive smooth
7	Simultaneous reflection and illumination estimation (SRIE) [20]	2016	The image reflection and illumination were estimated based on a weighted variational model to enhance images	Long processing time on complicated computation
8	Bright channel prior based on Retinex (BCPR) [21]	2019	The reflection and illumination maps were decomposed using a hybrid L2-Lp variational model for image enhancement	Long processing time on complicated computation
9	Image decomposition by semi-decoupling (SDD) [22]	2020	The Retinex image decomposition was performed using a semi-decoupled method for low-light image enhancement	Long processing time on complicated computation
10	LLnet [23]	2017	A deep autoencoder model was built by a stacked sparse denoising autoencoder to enhance low-light images	Paired training with different illumination images
11	KinD [24]	2019	The network was established based on the Retinex theory for dark image processing	Paired training with different illumination images
12	RDGAN [25]	2019	A generative adversarial network was integrated with the Retinex decomposition algorithm for image enhancement	Paired training with different illumination images
13	EnlightenGAN [26]	2019	An unsupervised generative adversarial network was proposed by utilizing a global-local discriminator structure and self-regularized perceptual loss function	Long processing time on complicated computation

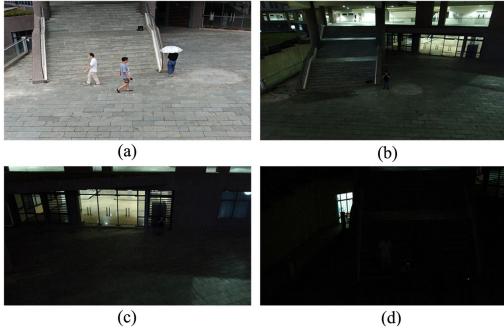


Fig. 1. UAV images captured under different illumination conditions. (a) Daytime. (b) Night-time with bright lights. (c) Night-time with weak lights. (d) Night-time with dim lights.

and sharpening are accomplished. A comparative experiment is carried out with the presented method and other recent image enhancement approaches. The image quality is measured, and the pedestrian detection accuracy is used to evaluate the performance of the proposed method.

The rest of this article is organized as follows. In Section II, the characteristics of UAV images captured at night are described. Section III details the proposed low-illumination image enhancement method that is a combination of the HTC and BM3D algorithms. Section IV describes the comparative experiment and the results are evaluated and discussed. Finally, Section V concludes this article.

II. UAV IMAGES CAPTURED AT NIGHT

The Mavic 2, a DJI UAV [27], is employed to detect pedestrians at night. The on-board camera can acquire images with

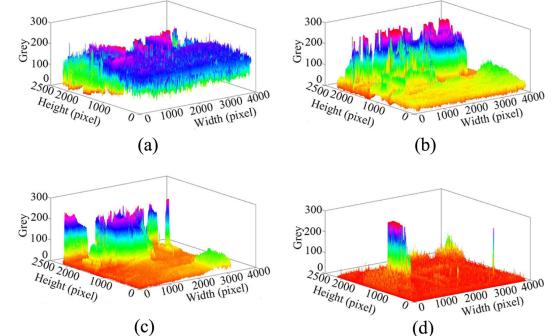


Fig. 2. Gray distribution of images shown in Fig. 1. (a) Gray distribution of Fig. 1(a). (b) Gray distribution of Fig. 1(b). (c) Gray distribution of Fig. 1(c). (d) Gray distribution of Fig. 1(d).

high resolution of 48 million pixels and 4K videos. Fig. 1 shows the UAV images captured under different illumination conditions, and their gray distributions are shown in Fig. 2. It can be seen that the gray values of the daytime image distribute more homogeneous than the other three images captured at night. Object detection and image segmentation algorithms [28] can be utilized to detect target objects in daytime images. However, in low illumination, the small gray gradient between the pedestrian and background makes it difficult for reliable detection. The image processing system may miss or misjudge the detected targets. This is a key issue that needs to be solved for night-time UAV surveillance.

III. LOW-ILLUMINATION IMAGE ENHANCEMENT FOR VISUAL OBJECT DETECTION AT NIGHT

In order to improve the gray contrast of pedestrian and image background for object detection at night, this article is focused

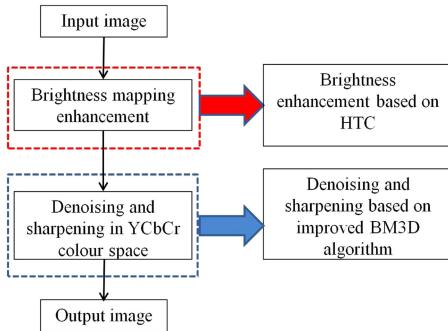


Fig. 3. Flowchart of the proposed image enhancement method.

on developing an image enhancement method. The flowchart of the proposed method is displayed in Fig. 3. First, image brightness is enhanced using HTC. Second, the BM3D algorithm is formulated by integrating an unsharp filter for image denoising and sharpening in the YCbCr color space.

A. Image Brightness Enhancement Based on the HTC

The HTC is a monotone increasing function in the range of $(-1, +1)$, which can be used for luminance mapping when image intensity values are normalized in $[0, 1]$ [29]. When an image is processed, the smaller image intensity values are increased more than the larger ones. Therefore, HTC is employed to enhance the low-illumination of the UAV images.

For an RGB image, let its pixel values be $I(x, y) = \{R(x, y), G(x, y), B(x, y)\}$. The red, green, and blue color components are represented as R , G , and B , respectively, and they are normalized to $[0, 1]$. The coordinate of pixel location is (x, y) . A hyperbolic tangent function

$$\tanh(\phi) = \frac{1 - \exp(-2\phi)}{1 + \exp(-2\phi)} \quad (1)$$

gives the nonlinear mapping of a color image, where $\phi = kI(x, y)$, and k is a scale factor. Its value is determined according to the image brightness to make sure that a higher brightness can be obtained.

In order to restrain the highlight area enhancement, the brightness of the output images is weighted as

$$I_\omega(x, y) = \omega I(x, y) + (1 - \omega) \tanh(kI(x, y)) \quad (2)$$

where ω is a weighting coefficient, which is assigned as the average of the RGB channel values to maintain the image intensity distribution.

The enhancement should cover all of the pixel gray scales from 0 to 1 to make sure that the image is globally enhanced. To this end, the weighted image is stretched by [30]

$$I_s(x, y) = \frac{I_\omega(x, y) - I_\omega(x, y)_{\min}}{I_\omega(x, y)_{\max} - I_\omega(x, y)_{\min}} \quad (3)$$

where $I_\omega(x, y)_{\max}$ and $I_\omega(x, y)_{\min}$ are the maximum and minimum of image intensity values, respectively.

It can be seen from (2) that the image enhancement results are influenced by the scale factor k . A suitable parameter is key to acquire a satisfactory processed image. In this work, the image

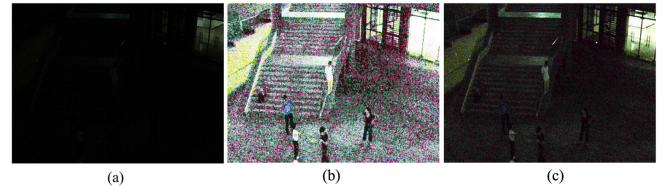


Fig. 4. Low-illumination image and its brightness enhancement results. (a) Low-illumination image. (b) Histogram equalization. (c) HTC-based enhancement.

entropy, as an indicator characterizing the image information, is considered as a measure of the enhanced image quality. The larger entropy indicates that the image carries more information and the image is sharper. The image entropy is defined as

$$J = - \sum_{i=0}^{255} p(i) \log_2 p(i) \quad (4)$$

where $p(i)$ is the proportion of the i th image intensity. Equation (4) is applied to calculate the entropy of $I_s(x, y)$ with different k . When the entropy is the largest, the corresponding k is the optimum value and *vice versa*. Accordingly, the optimum k is adaptively obtained by an iterative search based on the image entropy.

Fig. 4(a) shows an example of a low-illumination image. The image brightness is 0.0154. Thus, (k_1, k_2) are originally set to $(5.9, 9.9)$. It can be seen that the pedestrian is difficult to be seen visually. After histogram equalization, as shown in Fig. 4(b), the image brightness is significantly improved. However, a lot of artefact noises are produced. Fig. 4(c) shows the enhancement result by the HTC processing. Fig. 4(c) shows the advantages of obtaining sufficient enhancement using the HTC algorithm and fewer artefacts than that of the histogram equalization processing.

B. Hybrid Unsharp Filtering and BM3D for Image Denoising and Sharpening

In order to remove the artefacts to improve the image quality, image denoising is implemented after the brightness enhancement. Many kinds of image transformation and classification methods were reported to reduce noises, such as wavelet transform (WT), discrete cosine transform (DCT), contourlet transform (CT), principle component analysis (PCA), and the k-singular-value decomposition (K-SVD) [31]–[33]. However, the WT is generally used to remove noises that their frequencies are known. The DCT is sometimes confused about the background information and noise. The CT cannot realize image translation invariance and may lead to spectrum overlap. The denoising effectiveness of the PCA method is influenced by the noise distribution. The K-SVD solution needs large amounts of computation. Therefore, these mentioned methods are not suitable to process, in real time, the night-time UAV images with random noise.

The BM3D-based image denoising method sparsely represents the images in their transform domain and then enhances the

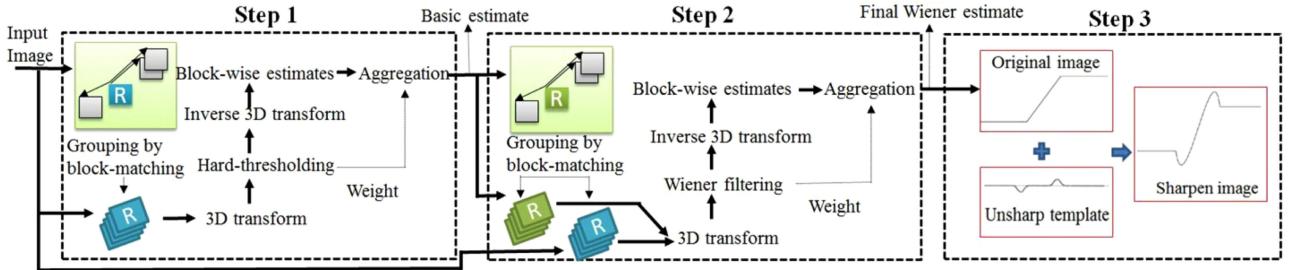


Fig. 5. Flowchart of the hybrid unsharp filter and BM3D-based denoising algorithm.

interested regions [34]. Similar image blocks are matched, and the image filtering in the 3-D transform domain is performed. Denoising based on block matching can retain some image detail information; hence, it is widely accepted for image denoising in many applications [35]. In order to recover image colors during the denoising progress, the RGB images are converted to the YCbCr color space for BM3D processing. In this space, the Y channel represents the brightness component of the image, and the Cb and Cr channels represent different ranges of color components.

In general, image noise is carried in the brightness component, and Cb and Cr channels carry color information. Accordingly, image denoising is focused on processing the Y channel information, and the Cb and Cr information is maintained without processing. Therefore, the processing efficiency is increased, and the detail color information is maintained as compared with the RGB image denoising.

The flowchart of the BM3D-based image denoising is displayed in Fig. 5. Three stages are involved, and they are the basic estimation, Wiener filtering, and image sharpening.

Step 1 (Basic estimation): The input image is divided into small blocks to be searched and matched. The 2-D matched blocks are overlapped with the input image to form a 3-D image set. Then, 3-D transform is conducted to obtain a sparse image group. Afterward, image denoising is performed using a hard-thresholding filter. Finally, the sparse image information is returned to the image domain by inverse 3-D transform, and the image blockwise estimates are obtained. Due to the uncertainties in the matching of similar blocks, the estimation results are corrected by aggregating all the blockwise estimates using appropriate weights.

During the hard-threshold filtering process, a threshold is preset according to the image acquisition conditions (2.7 in this work). The pixels whose gray values are lower than the preset threshold are regarded as noise, and their gray values are set to zero. The other pixels remain the same.

For image block matching, the Y channel image $Y(x, y)$ is divided into small blocks with the same size of $W \times W$. The image block with the top-left coordinate (x, y) is expressed as Y_x , and its reference image block is denoted as Y_{xr} . The L_2 norm is used to determine if the Y_x block is similar to the Y_{xr} block or not. The L_2 distance between Y_x and Y_{xr} is defined as

$$d(Y_{xr}, Y_x) = \frac{\|Y_{xr} - Y_x\|_2^2}{W^2}. \quad (5)$$

It can be seen that the small normalization factor W would give a larger distance, leading to overlapped blocks Y_x and Y_{xr} and producing matching errors. To solve this problem, the matching blocks are first normalized by a 2-D linear transformation. Then, the distance is thus modified as

$$d(Y_{xr}, Y_x) = \frac{\|\mathbf{T}_{2D}(Y_{xr}) - \mathbf{T}_{2D}(Y_x)\|_2^2}{W^2} \quad (6)$$

where $\mathbf{T}_{2D}(\cdot)$ denotes normalization with a 2-D linear transformation. When the distance is less than the matching threshold τ_{match} , the Y_x block is regarded as similar to the reference block Y_{xr} . The coordinate of the top-left corners of block Y_{xr} can be inferred from equation (6), that is,

$$S_{xr} = \{(x, y) \in (X, Y) | d(Y_{xr}, Y_x) \leq \tau_{\text{match}}\} \quad (7)$$

where (X, Y) is the coordinate of the input image.

In general, one reference block has more than one similar block, where the number of matched blocks is denoted as $|S_{xr}|$. The matched blocks are merged into an image block $Y_{S_{xr}}$ in three dimensions $W \times W \times |S_{xr}|$. Meanwhile, the selected image block Y_x can also be matched to other reference blocks. Hence, the same image block can be merged in different block sets.

The overlapped block sets are processed with a normalized 3-D linear transformation. The 3-D transform can be regarded as the superposition sum of 1-D and 2-D transforms that are realized by the Haar transform and the Bior transform, respectively [34]. The correlation of the image information and its transformed result is calculated and employed to sparsely represent the image block set. Then, image noise is removed by hard-thresholding filtering. After that, inverse 3-D transformation is performed to obtain the block estimates of the 3-D image set, that is,

$$\widehat{Y}_{S_{xr}} = \mathbf{T}_{3D}^{-1}(\Upsilon(\mathbf{T}_{3D}(Y_{S_{xr}}))) \quad (8)$$

where $\widehat{Y}_{S_{xr}}$ is the estimation of $Y_{S_{xr}}$.

Finally, the basic estimation of the input image can be obtained by aggregating all the estimates of reference image blocks, that is

$$\widehat{Y}_{\text{basic}}(x) = \frac{\sum_{xr \in X} \sum_{x_m \in S_{xr}} w_{xr} \widehat{Y}_{x_m}(x)}{\sum_{xr \in X} \sum_{x_m \in S_{xr}} w_{xr} \chi_{x_m}(x)} \quad \forall x \in X \quad (9)$$

where $\chi_{x_m} : X \rightarrow \{0, 1\}$ is the characteristic function of a block located at x_m , $\widehat{Y}_{x_m}(x)$ is the blockwise estimate of a block located at x_m , and w_{xr} is a weighted factor. When $\widehat{Y}_{x_m}(x) = 0$, $\chi_{x_m} = 0$; otherwise, $\chi_{x_m} = 1$.

The weighting operation is used to improve the estimate accuracy, which can be calculated by

$$w_{xr} = \begin{cases} \frac{1}{\sigma^2 N_{xr}}, & N_{xr} \geq 1 \\ 1, & N_{xr} = 0 \end{cases} \quad (10)$$

where σ is the standard deviation of block noises and N_{xr} is the number of nonzero elements after denoising by hard-thresholding.

For the images captured by the Mavic 2 DJI UAV, the standard deviation of image noise is generally less than 40. According to [34], the image block size ($W \times W$) is set to 8×8 pixels. In order to reduce border effects acting on the divided blocks, a Kaiser window is used as a part of the weight, and the window size is 8×8 pixels accordingly.

Step 2 (Wiener filtering): It can be seen from Fig. 5 that Step 2 is similar to Step 1, whereas the basic estimation result acts as the input for image block matching. The locations of similar blocks are searched by block matching with the basic estimates. This can improve the block matching accuracy because noise is reduced. Then, the basic estimate blocks and the input image blocks are merged to produce two 3-D image block sets. After the 3-D transform, the Wiener filter is used to remove image noise instead of hard-threshold filtering. This is because the Wiener filter can ensure the minimum mean square error between the filtered result and the ideal output image. Accordingly, Step 1 is performed to acquire a basic estimation that is regarded as the ideal output image and used to determine the Wiener coefficient (w_w)

$$w_w = \frac{\left| T_{3D}(\hat{Y}_{\text{basic}}) \right|^2}{\left| T_{3D}(\hat{Y}_{\text{basic}}) \right|^2 + \sigma^2}. \quad (11)$$

The input image with noise is converted by 3-D transform and then is multiplied by the Wiener coefficient. By doing so, the image noise is filtered. After the inverse 3-D transform, the Wiener estimates of similar blocks can be obtained as

$$\hat{Y}_{\text{wiener}} = T_{3D}^{-1}(w_w T_{3D}(Y_{S_{xr}})). \quad (12)$$

For the weight calculation in the Aggregation stage (Step 2), it can be defined as

$$w_{xr}^{(\text{wiener})} = \sigma^{-2} \|w_w\|_2^{-2}. \quad (13)$$

The final Wiener estimate can be obtained as

$$\hat{Y}_{\text{final}}(x) = \frac{\sum_{xr \in X} \sum_{x_m \in S_{xr}} w_{xr}^{(\text{wiener})} \hat{Y}_{x_m}(x)}{\sum_{xr \in X} \sum_{x_m \in S_{xr}} w_{xr}^{(\text{wiener})} \chi_{x_m}(x)} \quad \forall x \in X. \quad (14)$$

The 3D transform in Step 2 is different from that in Step 1, which is performed by the Haar transform and the DCT.

Step 3 (Image sharpening): In Steps 1 and 2, image noises are removed using the hard-threshold filter and the Wiener filter, which may cause some image signal loss and smooth the image edges. Correspondingly, an unsharp filter is embedded in the BM3D process for image sharpening to improve the image contrast. This process is the so-called unsharp masking. The unsharp template is the difference of the input image and

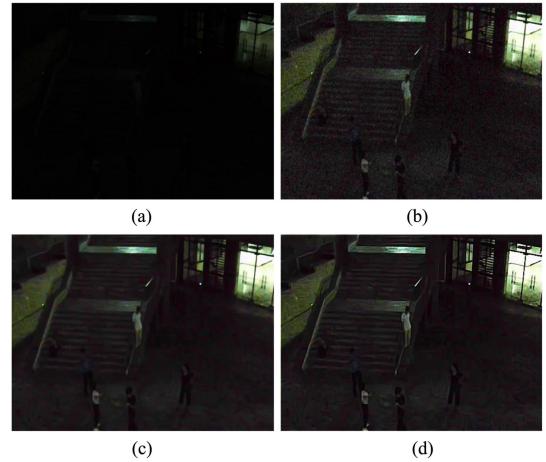


Fig. 6. Example of night-time image enhancement. (a) Input image. (b) HTC-based image enhancement. (c) BM3D-based image denoising. (d) Image sharpen.

the blurred image resulting from the Wiener filter. The high-frequency component, which is used for sharpening, is obtained by subtracting the blurred image from the input image. This process can be described as [36]

$$\hat{Y}_{\text{sharpen}}(x, y) = \hat{Y}_{\text{final}}(x, y) + \lambda S(x, y) \quad (15)$$

where λ is a scaling factor and $\lambda = 1$ for unsharp masking; $S(x, y)$ is the correction signal and is obtained by convolutional calculation with the Laplacian operator

$$S(x, y) = \begin{bmatrix} -h_1 & -h_2 & -h_1 \\ -h_2 & 1 + \frac{4}{\alpha+1} & -h_2 \\ -h_1 & -h_2 & -h_1 \end{bmatrix} * \hat{Y}_{\text{final}}(x, y)$$

$$h_1 = \frac{\alpha}{\alpha+1}, \quad h_2 = \frac{1-\alpha}{\alpha+1} \quad (16)$$

where α is an adjustment parameter of the Laplacian template in range of $[0, 1]$. The α value is dependent on the brightness of the night-time images. For the UAV images and the working conditions, the α is set to 0.8.

The sharpened Y channel image is then combined with the Cb and Cr channel components and reconverted to the RGB domain to obtain the final enhanced image. An example of a night-time image enhancement result is shown in Fig. 6.

IV. EXPERIMENT AND DISCUSSION

Comparison experiments are carried out to evaluate the performance, efficiency and accuracy, of the proposed low-illumination image enhancement method. The compared algorithms have been mentioned in the literature review, including AAQR [14], AGCCPF [15], MF [16], MSRCR [17], PLM [18], LIME [19], SRIE [20], BCPR [21], SDD [22], LLNet [23], KinD [24], RDGAN [25], and EnlightenGAN [26].

Five hundred pedestrian images are captured under bright, weak, and dim illuminations from street sides, roads, and campuses at night. The image resolution is 480×360 pixels. The program is coded in the MATLAB 2018a platform. The computer system is Windows 10, 64 bit, with 8-GB RAM; the

TABLE II
PROCESSING EFFICIENCIES OF DIFFERENT
IMAGE ENHANCEMENT ALGORITHMS

Algorithm	RTPF	FPS	Algorithm	RTPF	FPS
Proposed	0.26s	3.85	SRIE [20]	10.93s	0.09
AAQR [14]	0.29s	3.45	BCPR [21]	8.15s	0.12
AGCCPF [15]	0.14s	7.14	SDD [22]	6.52s	0.15
MF [16]	0.68s	1.47	LLNet [23]	2.89s	0.35
MSRRCR [17]	0.57s	1.75	KinD [24]	1.21s	0.83
PLM [18]	0.13s	7.69	RDGAN [25]	0.50s	2.00
LIME [19]	0.25s	4.00	EnlightenGAN [26]	0.80s	1.25

CPU is Intel Core i5-4210M, 2.6 GHz, and a NVIDIA 840-MB GPU.

A. Comparative Experiment for Efficiency Evaluation

The captured night-time images are processed by the mentioned fourteen image enhancement algorithms, and the efficiencies are listed in Table II. RTPF is the running time per image. FPS is frame per second. It shows that the PLM has the highest processing efficiency (0.13 s of RTPF and 7.69 FPS). The AGCCPF also shows satisfactory efficiency with 0.14 s of RTPF and 7.14 FPS. The proposed method has similar efficiency (3.85 FPS) to the AAQR (3.45 FPS) and the LIME (4.00 FPS). Other methods have lower efficiencies that are less than 2.00 FPS, indicating that they are not suitable for UAV visual detection due to its real-time requirement.

It needs to be mentioned that the processing efficiencies of deep learning methods can be improved by integrating a high-configuration GPU such as NVIDIA TITAN X GPU, NVIDIA 2080Ti GPU, and NVIDIA GTX 1080Ti GPU. However, these GPU products are more expensive than the consumer-grade UAVs such as the DJI MAVIC 2 [27]. High configuration will cost more than double, which is unconducive to promote the UAV monitoring technology in industrial applications.

B. Comparative Experiment for Accuracy Evaluation

According to the efficiency analysis, deep learning methods have lower efficiencies compared with traditional image enhancement methods. Hence, this section only presents the accuracy evaluation conducted with the AAQR, the AGCCPF, the LIME, the PLM, and the proposed method.

1) *Qualitative Analysis*: Fig. 7(a) shows eight example images captured in different illumination conditions. Fig. 7(b)–(f) shows the corresponding enhancement results by the AAQR, the AGCCPF, the LIME, the PLM, and the proposed method, respectively. However, the images enhanced by the AAQR, AGCCPF, and PLM methods show artefact noises, shown as Fig. 7(b), (c), and (e). Moreover, Fig. 7(b) and (d) shows excessive enhancement caused by AAQR and LIME processing. In contrast, the enhanced images in Fig. 7(f) processed by the proposed method contain less noise and are smoother than that of other methods.

2) *Quantitative Analysis*: Image quality assessment is conducted. Due to the difficulty of acquiring reference images, no-reference evaluation indicators, including the Minkowski distance measurement (MDM) [37], the contrast-changed image quality measure (CEIQ) [38], and the expected measure of

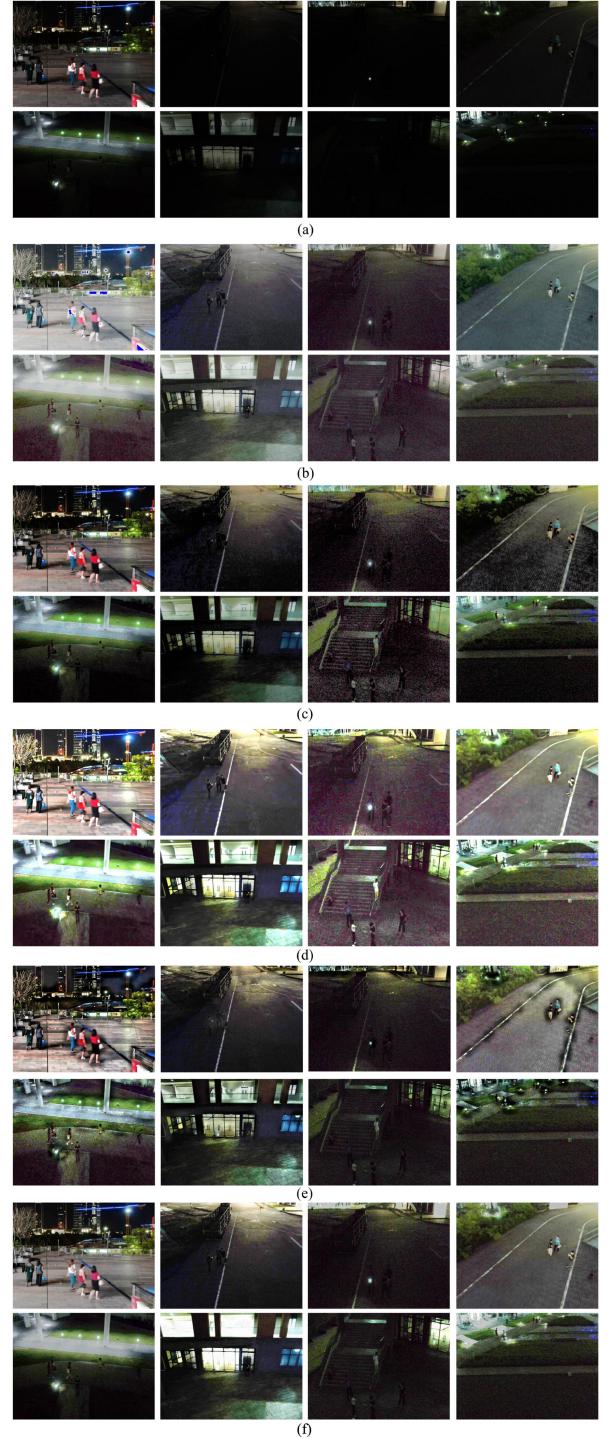


Fig. 7. Eight example low-illumination images and their corresponding enhancement results obtained by different methods. (a) Eight example images captured in different illumination conditions. (b) AAQR. (c) AGCCPF. (d) LIME. (e) PLM. (f) Proposed method.

enhancement by gradient (EMEG) [39], are used to measure the enhanced UAV images. The MDM, CEIQ, and EMEG of 500 enhanced images are displayed in Fig. 8. It can be seen that the MDM (0.975) of the proposed approach is the highest compared with other methods. However, the LIME obtains

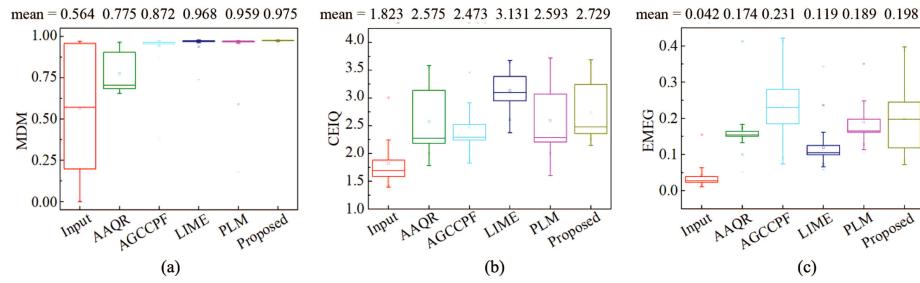


Fig. 8. Boxplot of MDM, CEIQ, and EMEG of 500 enhanced images processed by different methods. (a) MDM. (b) CEIQ. (c) EMEG.

the highest CEIQ (3.131) and the AGCCPF has the highest EMEG (0.231).

As mentioned above, AGCCPF processing produces artefact noise. Although the EMEG of the AGCCPF is the highest, its MDM and CEIQ values are both lower than the proposed method. LIME processing causes excessive enhancement that can significantly improve image contrast. However, the excessive enhancement results are not the expected output. This phenomenon indicates that different measure indexes reflect different image characteristics. The performance of image enhancement cannot be evaluated using only one indicator.

3) Pedestrian Detection for Further Evaluation: The purpose of image enhancement is to improve the accuracy of UAV pedestrian detection at night. Hence, pedestrian detection is performed to further evaluate the performance of the proposed image enhancement method. In this work, the YOLOv3 [40] is considered as the pedestrian detection model because it is more suitable for object detection in various scenes and has higher detection accuracy than the SSD [41].

The YOLOv3 model is trained with a public dataset of pedestrian images. Then, the enhanced images, shown in Fig. 7, are used as testing samples for pedestrian detection. Fig. 9 shows the detection results. The number marked above the red box is the confidence coefficient (Cc) of pedestrian detection. It can be seen from Fig. 9(a) that some pedestrians are not detected in the weak and dim illumination conditions. After enhancement by the AAQR, AGCCPF, LIME, PLM, and the proposed method, the detection accuracy is improved, as shown in Fig. 9(b)–(f). Moreover, the detection Ccs of enhanced images are higher than the input images.

Fig. 9(d) shows that some target objects are misdetected by LIME despite that its CEIQ is the highest. Meanwhile, missing detection also occurs from the AGCCPF and the PLM enhancement images, as shown in Fig. 9(c) and (e). It seems that the AAQR and the proposed method can provide more satisfactory detection results. For further analysis, the F-measure (an accuracy index) [42] and the Cc of all detection results from the enhanced images are calculated and shown in Fig. 10. The boxplots show that the F-measure (0.907) and Cc (0.840) of the pedestrian detection from the enhanced images processed by the proposed method are both the highest. The proposed method also shows a higher robust detection capability than other methods. These observations indicate that the proposed

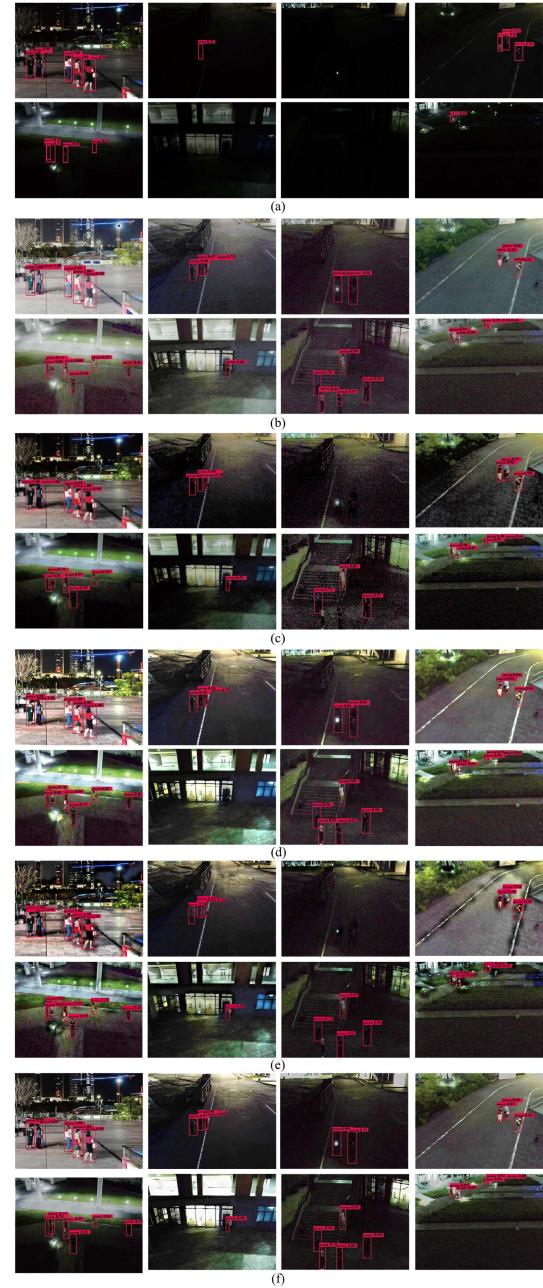


Fig. 9. YOLOv3-based pedestrian detection of the images shown in Fig. 7. (a) Input images. (b) AAQR. (c) AGCCPF. (d) LIME. (e) PLM. (f) Proposed method.

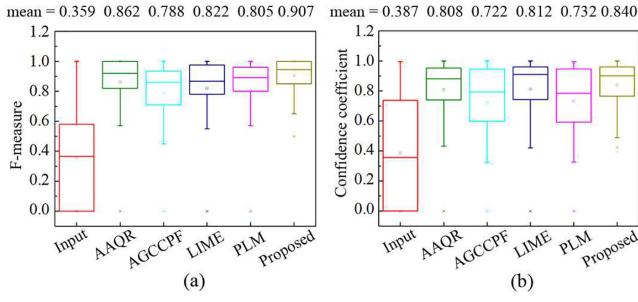


Fig. 10. Boxplot of F-measure and Cc of 500 enhanced images by different methods. (a) F-measure. (b) Confident coefficient.

TABLE III
AVERAGE MDM, ECIQ, AND EMEG VALUES OF 200 ExDARK IMAGE SAMPLES PROCESSED BY DIFFERENT METHODS

Indicator	Input	AAQR	AGCCPF	LIME	PLM	Proposed
MDM	0.817	0.861	0.942	0.962	0.970	0.973
CEIQ	2.153	2.881	2.689	3.174	3.058	3.176
EMEG	0.071	0.164	0.157	0.125	0.236	0.211

low-illumination image enhancement method is able to improve UAV visual detection accuracy at night.

C. Generalization Ability Evaluation

In order to evaluate the generalization ability of the proposed image enhancement method, two hundred low-illumination images were selected from the Exclusively Dark database (ExDark) [43] to further evaluate the performance. The comparative experiment was also conducted with the AAQR, the AGCCPF, the LIME, the PLM, and the proposed method.

Table III shows the quantitative results of the image quality assessment indicators (MDM, EMEG, and CEIQ). It can be seen that the average MDM, CEIQ, and EMEG values of the input EXDark image samples are 0.817, 2.153, and 0.071, respectively. After enhancement, the proposed method acquired the highest MDM (0.973) and the highest CEIQ (3.176), and the second highest EMEG (0.211), which are similar to that shown in Fig. 8. This indicates that the proposed method can be generalized to new images.

Fig. 11 shows four example low-illumination images selected from the ExDark dataset and the pedestrian detection results after enhancement by different methods. It can be seen that the brightness of all input images is increased. The image contrasts and the details are also enhanced at the same time using different image enhancement methods. In contrast, the images enhanced by the proposed method are clearer with less noise compared with other approaches.

We need to mention that the open-source database ExDark is acquired using different kinds of cameras under various scenes. The proportion of people in daily-life photos is generally higher than the UAV images. The influence of noise on the daily-life photos is also less than the UAV images. Accordingly, pedestrian detection from the ExDark images is much easier than that from the UAV images. All the methods of AAQR, AGCCPF, LIME, PLM, and the proposed method obtain satisfactory accuracies of pedestrian detection from the ExDark image samples, as shown in Fig. 11.

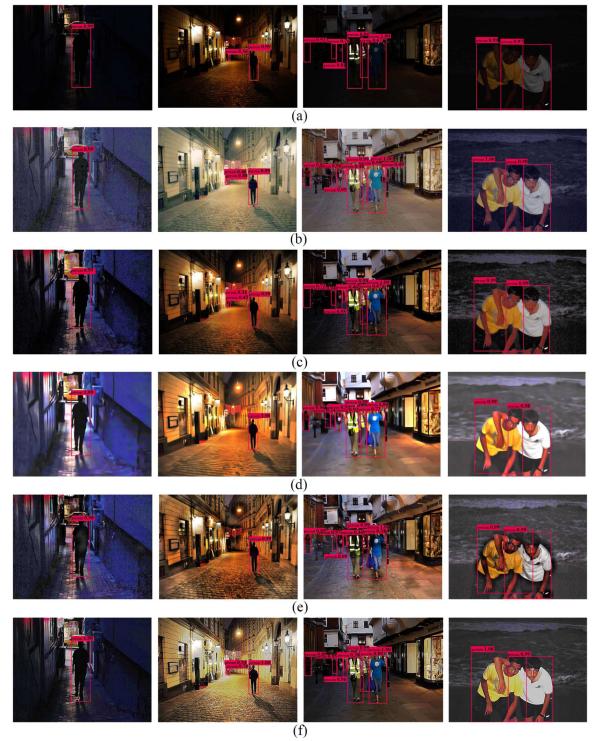


Fig. 11. Four example low-illumination image samples selected from the ExDark dataset and the pedestrian detection using different methods. (a) Input images. (b) AAQR. (c) AGCCPF. (d) LIME. (e) PLM. (f) Proposed.

V. CONCLUSION

Aiming at improving object detection accuracy of UAV cruise at night, in this article a hybrid low-illumination image enhancement method was developed by integrating the optimal HTC and the improved BM3D algorithms. The following conclusions are drawn.

- 1) The pedestrian detection accuracy and Cc after image enhancement by the proposed method reach 0.907 and 0.840, which were both higher than that of other compared methods.
- 2) Experimental results also showed that the proposed method has higher robustness in object detection than other methods. This indicated that the proposed low-illumination image enhancement method is more suitable to be applied in UAV cruise detection systems.
- 3) The processing rate of the proposed method was 3.85 FPS, which was lower than that of AGCCPF (7.14 FPS), LIME (4.00 FPS), and PLM (7.69 FPS).

Future work will focus on algorithm optimization to improve processing efficiency for real-time pedestrian detection. Furthermore, the proposed method needs to adaptively obtain the image processing parameters to deal with different detection issues with varying illumination conditions.

REFERENCES

- [1] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "UAV-enabled intelligent transportation systems for the smart

- city: Applications and challenges," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22–28, Mar. 2017.
- [2] B. Zhang, C. H. Liu, J. Tang, Z. Xu, J. Ma, and W. Wang, "Learning-based energy-efficient data collection by unmanned vehicles in smart cities," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1666–1676, Apr. 2017.
- [3] X. Jiang, J. Sun, C. Li, and H. Ding, "Video image defogging recognition based on recurrent neural network," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3281–3288, Jul. 2018.
- [4] Z. Ren, H. K.-H. So, and E. Y. Lam, "Fringe pattern improvement and super-resolution using deep learning in digital holography," *IEEE Trans. Ind. Informat.*, vol. 15, no. 11, pp. 6179–6186, Nov. 2019.
- [5] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, 2018.
- [6] S. Ko, S. Yu, W. Kang, C. Park, S. Lee, and J. Paik, "Artifact-free low-light video enhancement using temporal similarity and guide map," *IEEE Trans. Ind. Electron.*, vol. 64, no. 8, pp. 6392–6401, Aug. 2017.
- [7] C. Zhao, G. Zhao, and X. Jia, "Hyperspectral image unmixing based on fast kernel archetypal analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 331–346, Jan. 2017.
- [8] Y. Park and I. S. Kweon, "Ambiguous surface defect image classification of AMOLED displays in smartphones," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 597–607, Apr. 2016.
- [9] Y. Peng, T. Wu, S. Wang, N. Kwok, and Z. Peng, "Motion-blurred particle image restoration for on-line wear monitoring," *Sensors*, vol. 15, no. 4, pp. 8173–8191, 2015.
- [10] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, 2018.
- [11] G. Bianco and L. Neumann, "A fast enhancing method for non-uniformly illuminated underwater images," in *Proc. OCEANS 2017 Conf.*, Anchorage, AK, USA, 2017, pp. 1–6.
- [12] L. Liu, Z. Jia, J. Yang, and N. Kasabov, "A medical image enhancement method using adaptive thresholding in NSCT domain combined unsharp masking," *Int. J. Imag. Syst. Technol.*, vol. 25, no. 3, pp. 199–205, 2015.
- [13] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Apr. 2017.
- [14] J. Shen *et al.*, "Nighttime driving safety improvement via image enhancement for driver face detection," *IEEE Access*, vol. 6, pp. 45625–45634, 2018.
- [15] B. Gupta and M. Tiwari, "Minimum mean brightness error contrast enhancement of color images using adaptive gamma correction with color preserving framework," *Optik—Int. J. Light Electron Opt.*, vol. 127, no. 4, pp. 1671–1676, 2016.
- [16] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, 2016.
- [17] B. Jiang, G. A. Woodell, and D. J. Jobson, "Novel multi-scale retinex with color restoration on graphics processing unit," *J. Real-Time Image Process.*, vol. 10, no. 2, pp. 239–253, 2015.
- [18] S. Y. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 28–37, Jan. 2019.
- [19] X. Guo, "LIME: A method for low-light image enhancement," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 87–91.
- [20] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2782–2790, 2016.
- [21] G. Fu, L. Duan, and C. Xiao, "A hybrid L_2 - L_p variational model for single low-light image enhancement with bright channel prior," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1925–1929.
- [22] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE Trans. Multimedia*, to be published.
- [23] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, 2017.
- [24] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 1632–1640.
- [25] J. Wang, W. Tan, X. Niu, and B. Yan, "RDGAN: Retinex decomposition based adversarial learning for low-light enhancement," in *Proc. IEEE Int. Conf. Multimedia Expo.*, pp. 1186–1191, 2019.
- [26] Y. Jiang *et al.*, "Enlightengan: Deep light enhancement without paired supervision," 2019, *arXiv:1906.06972*.
- [27] Resource Description Document of Mavic 2, a DJI UAV, 2019. [Online]. Available: https://www.dji.com/cn/mavic-2?site=brandsite&from=landing_page#zoom
- [28] H. Yang, J. Wen, X. Wu, L. He, and S. G. Mumtaz, "An efficient edge artificial intelligence multi-pedestrian tracking method with rank constraint," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4178–4188, Jul. 2019.
- [29] S. Liu *et al.*, "Enhancement of low illumination images based on an optimal hyperbolic tangent profile," *Comput. Electr. Eng.*, vol. 70, pp. 538–550, 2018.
- [30] J. Zhang, J. Jia, A. Sheng, and K. Hirakawa, "Pixel binning for high dynamic range color image sensor using square sampling lattice," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2229–2241, May 2018.
- [31] F. Zhang, B. Du, L. Zhang, and L. Zhang, "Hierarchical feature learning with dropout k-means for hyperspectral image classification," *Neurocomputing*, vol. 187, pp. 75–82, 2016.
- [32] G. Yang, Z. Lu, J. Yang, and Y. Wang, "An adaptive contourlet HMM-PCNN model of sparse representation for image denoising," *IEEE Access*, vol. 7, pp. 88243–88253, 2019.
- [33] J. V. Manjón, P. Coupé, and A. Buades, "MRI noise estimation and denoising using non-local PCA," *Med. Image Anal.*, vol. 22, no. 1, pp. 35–47, 2015.
- [34] D. Kostadin, F. Alessandro, and E. Karen, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. Eur. Signal Process. Conf.*, 2007, pp. 145–149.
- [35] Z. Hua and Y. Zhou, "Design of image cipher using block-based scrambling and image filtering," *Inf. Sci.*, vol. 396, pp. 97–113, 2017.
- [36] S. Lin *et al.*, "Intensity and edge based adaptive unsharp masking filter for color image enhancement," *Optik*, vol. 127, no. 1, pp. 407–414, 2016.
- [37] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 518–523, Jun. 2018.
- [38] J. Yan, J. Li, and X. Fu, "No-reference quality assessment of contrast-distorted images using contrast enhancement," 2019, *arXiv:1904.08879*.
- [39] T. Celik, "Spatial entropy-based global and local image contrast enhancement," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5298–5308, Dec. 2014.
- [40] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [42] G. Cao, S. Ruan, Y. Peng, S. Huang, and N. Kwok, "Large-complex-surface defect detection by hybrid gradient threshold segmentation and image registration," *IEEE Access*, vol. 6, pp. 36235–36246, 2018.
- [43] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Comput. Vis. Image Understanding*, vol. 178, pp. 30–42, 2019.



Weijiang Wang received the B.E. degree in electrical engineering and automation from Huizhou University, Huizhou, China, in 2018. He is currently working toward the M.Sc. degree in control engineering with Shenzhen University, Shenzhen, China.

His current research interests include machine vision, image enhancement, and defect detection.



Yeping Peng (Member, IEEE) received the B.E. degree in mechanical design, manufacture, and automation from Harbin Engineering University, Harbin, China, in 2011, and the M.Sc. and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2014 and 2017, respectively.

She is currently an Associate Research Fellow with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China. Her main research interests include machine vision and image processing.



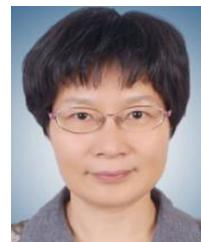
Guangzhong Cao (Senior Member, IEEE) received the B.E., M.Sc., and Ph.D. degrees in electrical engineering and mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 1989, 1992, and 1996, respectively.

He is currently a Professor and the Director of the Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, Shenzhen University, Shenzhen, China. His current research interests include control theory, machine vision, and information processing.



Ngaiming Kwok received the M.Phil. degree in control engineering from Hong Kong Polytechnic University, Hong Kong, in 1997, and the Ph.D. degree in mobile robotics from the University of Technology Sydney, Sydney, NSW, Australia, in 2007.

He is currently with the University of New South Wales, Sydney. His current research interests include image processing, intelligent computation, and automatic control.



Xiaoqin Guo received the B.E. and M.Sc. degrees in aerospace engineering from the Department of Aerospace Engineering, Northwestern Polytechnical University, Xi'an, China, in 1982 and 1987, respectively.

She is currently an Associate Professor with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China. Her main research interests include machine vision and image processing.