Simulation Generation Algorithm for Foggy Images in Natural Scenes

Jianping Liu(✉), Qing Ye, Shizhuo Qiu, and Yuze Liu

School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, Hunan, China
799372927@qq.com

Abstract. The foggy environment seriously affects the automatic inspection or cruise monitoring of outdoor equipment in the power system. Aiming at the lack of fog image data sets, a fog simulation image generation algorithm based on depth estimation was proposed. First of all, by unsupervised depth estimation model building the depth map of an outdoor clear picture. Then using feature fusion to refine the depth chart details. By setting the atmospheric extinction coefficient for transmittance figure. Dark channel method is used to estimate the atmospheric light value of the image. Finally, the fog simulation images are obtained based on the atmospheric scattering model. The experimental results show that the method improves the depth map effectively, and the generated fog simulation image is reliable. The average error rate of fog simulation is 6.2%, which solves the problem of excessive uneven fog edge, and the fog simulation image effect is very good in low visibility. Generating fog images with different visibility labels can solve the problem of lack of fog datasets.

Keywords: Fog Simulation · Depth Estimation · Automatic Inspection

1 Introduction

Under fog weather conditions, the scattering and absorption of light caused by a large number of tiny water droplets suspended in the atmosphere greatly reduce the visibility, resulting in lower image quality and affecting the accuracy of computer vision, such as automatic driving, object detection, robot inspection of power system equipment, etc. Therefore dehazing and sharpening of video images has become a key research content in computer vision. A large number of machine learning and deep learning dehazing algorithms need to be trained on foggy images. Due to the randomness of fog, it is difficult to collect image data sets under real fog scenes, and the number of existing natural foggy image datasets is difficult to meet the needs of research and training.

Based on Koschmieder's law [1], the influence of atmospheric medium on image information in foggy outdoor environment depends on the distance between scene information and image acquisition equipment, that is, scene depth. The authenticity of fog simulation image depends on the authenticity of the generated scene depth. Researchers

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 173–185, 2023.
https://doi.org/10.1007/978-3-031-31733-0_16

have used a variety of methods to obtain scene depth, including direct measurements using LiDar equipment and methods based on two images and videos. Methods in computer vision include binocular disparity depth estimation, optical flow method, monocular depth estimation, etc. In the aspect of scene depth estimation, monocular depth estimation based on unsupervised learning has become a hot research topic in this field because it does not rely on depth truth value during network training. Gary [2] proposed the use of stereo images to realize unsupervised monocular depth estimation without depth label, and its working principle is similar to autoencoder. Based on the above work, Godard [3] uses the consistency of left and right views to realize unsupervised depth prediction, further adding details and improving resolution. However, the above methods have limited ability to predict depth, and the prediction effect is not good for the scene with rich details. Chen [5] created a new dataset 'Depth in the Wild', which contained the relative Depth relationship between any image and random points in the image, and proposed an algorithm to estimate Depth by using the relative Depth relationship. The algorithm can restore the details of the image well, but the prediction depth needs to be improved.

In summary, the existing fog simulation methods are mainly based on the real information or binocular disparity data to obtain the depth map, but from the reality, it is difficult to obtain the real depth information and binocular image. Therefore, this paper conducts monocular depth estimation based on unsupervised learning of real scenes, uses feature fusion to refine the depth map details, and combines atmospheric scattering model to generate high-quality fog simulation images.

2 Atmospheric Scattering Model

Under fog conditions, there are a large number of suspended particles in the atmosphere to absorb and scatter light [6], which makes the imaging results of the detection system degrade. There are two main reasons: first, the reflected light of the target is absorbed and scattered by the suspended particles, resulting in the attenuation of brightness and contrast; Second, ambient light such as atmospheric light forms background light due to the scattering effect of medium, and is stronger than the target reflected light, which often makes the image blurred. Therefore, the light source of foggy image imaging is mainly the light that the target reflected light attenuates to the imaging system and the atmospheric light that the light source is scattered by particles. Based on this principle, the atmospheric scattering model widely used today is simplified, and its mathematical model expression is as follows.

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

$$t(x) = e^{-\beta d(x)} \tag{2}$$

where I(x) is the foggy image; J(x) is the clear image; A is the atmospheric light value at infinity; t(x) is the atmospheric transmittance map, formula (2) is the expression of the transmission map, where β is the atmospheric scattering coefficient, d(x) is the distance between the target and the camera, also known as scene depth, and the depth map is represented by thermal image quantification.

Based on the above atmospheric scattering model and its principle, the fog-free clear images in natural environment are simulated as foggy images. Under the premise that the atmospheric scattering coefficient is determined by the visibility, the estimation of depth map, transmittance map and atmospheric light value becomes the main content of the study.

3 Parameters Estimation

3.1 Estimation of Atmospheric Light Values

Since the clear original image does not have the influence of large pixel points caused by fog [7–9], its atmospheric light value can be directly estimated by using the pixel point of the brightest region in the image, which is generally the value of the sky region or infinity. However, the single brightest pixel in the image tends to appear on small white objects, such as vehicles and white floor tiles, and the atmospheric light value determines whether the fog color is normal after fog simulation. Therefore, the dark channel [11] method is used to estimate the atmospheric light value.

In the clear image without sky area, the dark channel approaches 0, while in the original image with sky area, the pixels close to the sky area do not approach 0. Therefore, the dark channel can be used to screen the sky area, and the three-channel pixels in this area are averaged, so as to obtain the estimated value of atmospheric light. If there is no sky area, the first 0.1% pixel in the dark channel corresponding to the brightest pixel in the original image is used as the estimate value of atmospheric light.

3.2 Depth Map Estimation

The self-supervised monocular depth estimation method is trained with video sequences or continuous images, which can input any monocular image and output a reliable depth map. In this paper, the *manydepth* [12] model is used to obtain image depth. Using the cost-volume [14] method in multi-view depth estimation [13], the model replaces the original binocular image with a pair of front and rear frame images, overcoming the scale ambiguity problem in depth estimation. The model is composed of pose estimation network, cost-volume construction and depth estimation network. The block diagram is shown in Fig. 1.

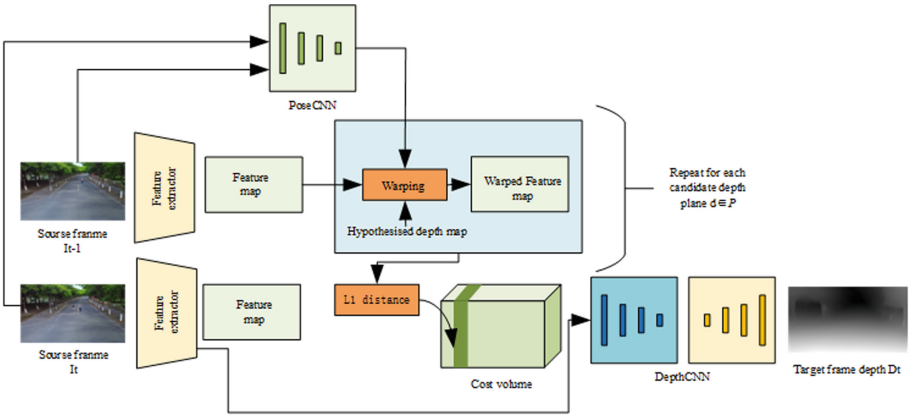

Fig. 1. Self-supervised monocular depth estimation framework

Pose Estimation Network

Here, two adjacent frames of images are used to estimate the pose of the camera $T_{t\to t+n}$ ($n = -1$), and the pose network is trained with the images of past adjacent frames. Then the pose is described as (3).

$$T_{t\to t+n} = \theta_{pose}(I_t, I_{t+n}) \tag{3}$$

Building a Cost-Volume

Cost-Volume is described as the difference in pixels of adjacent frames at different depths, but it is not the difference in images but the difference in features. According to the maximum depth d_{max} and minimum depth d_{min}, several planes P are divided in the vertical direction of the I_t optical axis. Then, $F_{t+n\to t,d}$, $d \in P$ is obtained by using the camera pose information and the internal reference matrix to transform the features of the source image. The cost-volume is obtained by making the absolute value of the difference between the source image features and the target image features. Finally, it is combined with the features of the target image to get the depth estimation map through the decoder.

Depth Estimation Network

The *manydepth* model is trained using multiple image sequences, so the depth estimation part is described as follows.

$$D_t = \theta_{depth}(I_t, I_{t-1}, \ldots, I_{t-n}) \tag{4}$$

Formula (4) shows that the model uses the image data of past frames as input to train and predict depth ($n = -1$). The estimation of camera pose $T_{t\to t+n}$ using the current estimated depth D_t and pose estimation network θ_{pose} is used to reconstruct the scene, but only using pixels from adjacent frames ($n = -1$), which is described as follows.

$$I_{t+n\to t} = I_{t+n}\langle pr(D_t, T_{t\to t+n}, K)\rangle \tag{5}$$

where <> represents the sampling operator and pr represents the 2D coordinates of the depth returned when *pr* reprojected into the camera of I_{t+n}. For each pixel, the loss of the best-matched source image is optimized by selecting the smallest pixel on the reconstruction loss *pe*, which can be expressed as follows.

$$L_p = \min_n pe(I_t, I_{t+n\to t}) \tag{6}$$

Refinement of the Depth Map

It is found that the resolution of the input image has the following two effects on the quality of the depth map. When the low resolution image is input, the generated depth map has good depth consistency, but the image detail performance is poor. When a high-resolution image is input, the predicted depth map has good image details but poor depth consistency performance. The results are shown in Fig. 2 and Fig. 3.

Fig. 2. High resolution image and its depth map

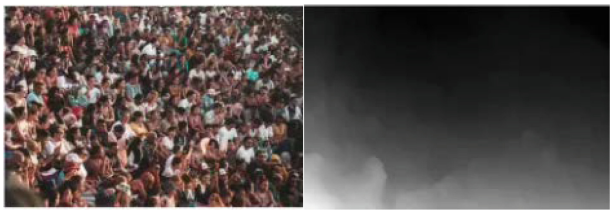

Fig. 3. Low image and its depth map

Figure 2 and Fig. 3 show that the characteristics of the monocular depth estimation network vary with the resolution of the input image, and the main reason is that the network receptive field is limited, and the size of the receptive field mainly relies on the network structure and the resolution of the training datasets, depth estimation depends on the image context clues, when image clues in farther than receptive field, the network cannot accept enough information.

To solve the above problem, we generate multiple depth maps for the individual images to be merged to achieve results with high frequency details with a consistent overall structure. The Pixel2pixel [16] structure with 10-layer U-net [15] is used as the generator of the merge network, which accepts two depth maps with different resolutions. After the two depth maps are resized, different sampling patches are selected according

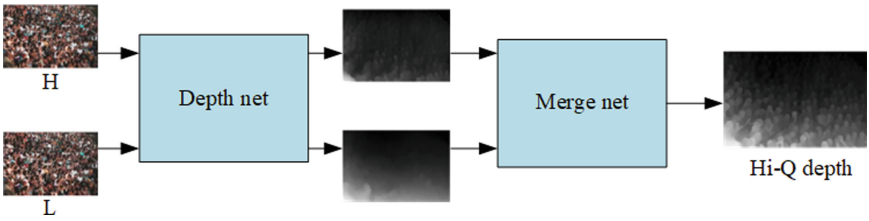

Fig. 4. Multi-scale fusion

to the set gradient, and then multi-scale fusion is performed through the Pixel2Pixel generator. Specific steps are shown in Fig. 4.

Figure 4 shows that by adding the Merge network, the depth image has the accurate and consistent scene structure of the boundary, which lays a foundation for the subsequent calculation of the transmittance map of fog images.

3.3 Transmission Map Estimation

According to formula (2), it can be concluded that the atmospheric extinction coefficient β is an important parameter to generate the transmittance map in addition to the depth map. According to *Koschmieder's* law, visibility is closely related to the atmospheric extinction coefficient, and the specific relationship is as follow.

$$V = -\frac{ln\sigma}{\beta} \tag{7}$$

where V is visible distance, namely visibility; β is the same as formula (2), is the atmospheric extinction coefficient, σ is the visual contrast threshold, generally 0.02 or 0.05, this paper takes 0.05; Based on this law, the corresponding β value can be generated according to the required visibility value, so that the estimated depth map can be used to calculate the pixels in the image one by one, and the corresponding transmittance map can be estimated.

Thus, the depth map, transmittance map and atmospheric light value needed for fog simulation are estimated. Foggy images with different visibility can be simulated according to the principle shown in formula (1): First use natural environment without clear mist image depth map estimation and atmospheric optical value estimation, then selected need visibility of atmospheric extinction coefficient is calculated, thus to estimate transmittance using atmospheric scattering model finally from the clear picture to the fog simulation, get visibility standard and the actual situation of image contrast. The process is shown in Fig. 5.

4 Results and Discussion

4.1 Improved Depth Map Image Contrast

The outdoor natural environment images captured by the camera were used for the experiment with a resolution of 1920 × 1080. Fog simulation was carried out before

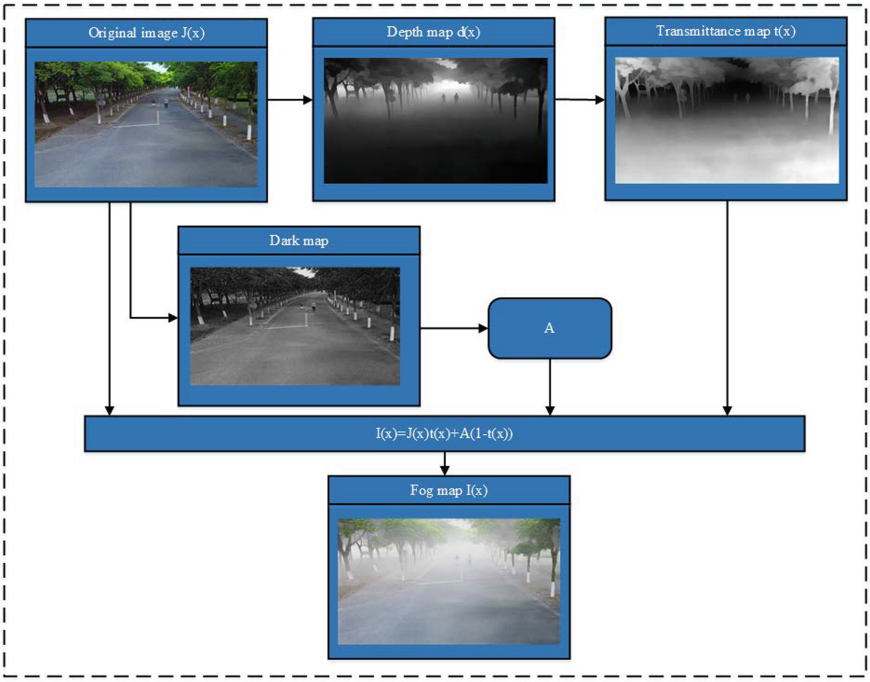

Fig. 5. Fog diagram simulation process

and after the improvement. As can be seen from Fig. 6, the perspective of the depth map and transmittance map without improvement is too smooth, the details of the image information are lacking, and the furthest depth of the depth map is not consistent with the actual situation, resulting in too uniform fog in the perspective of the fog map. There is no transition between the fog area and the non-fog area, which is inconsistent with the real foggy image. Multi-scale feature fusion is used to improve the depth map. First, the original image is clipped by line, and the original image and the clipped image with two different resolutions are respectively input into the depth estimation model, and the obtained results are merged to obtain the improved depth map. It can be seen from Fig. 7 that the details of the depth map and transmittance map are obvious, which are more real than Fig. 6. The simulated fog map also has obvious depth information, and there is a good transition between the fog area and the non-fog area, which is in line with the actual situation.

Fig. 6. Images without improvement. Clear image; Depth map; Transmission map; Fog simulation image

Fig. 7. Images with improvement. Clear image; Depth map; Transmission map; Fog simulation image

BRISQUE [17], NIQE [18] and PIQE [19] are selected to evaluate the quality of simulated images in accordance with human visual effects. The smaller the value, the better the quality and the more in line with human visual observation. As shown in Table 1, adding Merge network effectively improves fog map quality.

Table 1. Image quality evaluation

Evaluating indicator	BRISQUE	NIQE	PIQE
Original	27.7	2.9	47.3
Improvement	24.3	1.6	19.2

4.2 Foggy Images Under Different Scenes

In meteorology, the intensity of fog is described by visibility, which is the maximum distance the human eye can identify an object from a background, usually the sky near the horizon. In the following, fog ($V = 100$ m), medium fog ($V = 500$ m) and mist ($V = 800$ m) were simulated through images of three different scenes in 3D composite image, cruise map of power equipment and urban road map in Cityscape datasets, as shown in Fig. 8, Fig. 9 and Fig. 10 (Table 2).

Table 2. The relationship of visibility and fog concentration

Category	Dense fog	Big fog	Moderate fog	Mist
Visibility(km)	<0.2	0.2~0.5	0.5~1.0	1.0~5.0

Fig. 8. Simulation of 3D modeling diagram. Visibility = 100 m; Visibility at = 500 m; Visibility = 800 m

Fig. 9. Simulation of urban road map. Visibility = 100 m; Visibility = 500 m; Visibility = 800 m

Fig. 10. Simulation of cruise map. Visibility = 100 m; Visibility = 500 m; Visibility = 800 m

4.3 Quality Evaluation of Foggy Simulated Images

In order to verify the validity of the simulated foggy images, the simulated foggy images were compared with the real foggy images. Figure 12 shows the real fog images when the visibility is 100 m, 300 m, 500 m and 800 m. Figure 13 shows the images with visibility of 100 m, 300 m, 500 m and 800 m in fog simulation of the same scene without fog. Figure 11 is a clear image, which is used to simulate the above real foggy images. From the subjective evaluation of human vision, the fog degree of images is similar, the image features are calculated and evaluated, and the image quality is compared using the evaluation index of image sharpness without reference.

Fig. 11. A clear image of the same scene.

Fig. 12. Real fog. Visibility = 100 m; Visibility = 300 m; Visibility = 500 m; Visibility = 800 m.

Laplacian [20] gradient function, SMD function, SMD2 function [21] and information entropy [22] function, which are commonly used to evaluate image sharpness without reference, are used for quantitative evaluation. Laplacian gradient function adopts

Fig. 13. Simulated fog image. Visibility = 100 m; Visibility = 300 m; Visibility = 500 m; Visibility = 800 m.

Laplacian operator to extract horizontal and vertical gradient values respectively for convolution operation. The higher the gradient value, the clearer the image is. The SMD function is also used to accumulate the sum of the absolute values of the difference between the gray values of adjacent pixels in two directions as the function value. The higher the value, the clearer the image will be. The SMD2 function multiplicate the two gray differences in each pixel neighborhood and then accumulates pixel by pixel, which improves the former shortcoming. The operation of multiplication and division is used to expand useful information. The larger the value, the clearer the image will be. The information Entropy function measures the richness of image information. The larger the value of the function, the richer the information and the more detailed features. Although the above indicators have certain limitations, generally speaking, the closer the value between the foggy image and the simulated image, the more realistic the simulated image is (Table 3).

Table 3. Comparison of evaluation indicators

Visibility	Category	Laplacian	SMD(107)	SMD2(107)	Entropy
100 m	Real fog	80.5	0.335	0.966	3.968
100 m	Simulated fog	79.3	0.331	0.953	3.637
300 m	Real fog	140.3	0.732	1.334	4.212
300 m	Simulated fog	139.1	0.712	1.322	4.207
500 m	Real fog	160.7	0.882	1.912	4.334
500 m	Simulated fog	159.1	0.812	2.121	4.451
800 m	Real fog	189.6	0.901	2.874	4.673
800 m	Simulated fog	187.3	0.873	2.889	4.771

4.4 Generation of Foggy Datasets

Using the above fog image generation algorithm, 765 urban road images are selected from Cityscape datasets to generate a total of 7650 fog data sets with visibility labels of 50 m, 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 700 m, 800 m and 900 m. The data sets are real and reliable. It can be used in many machine vision fields such as image recognition and object detection. Part of the datasets are shown in Fig. 14.

Fig. 14. Datasets of simulated foggy image. Visibility = 50 m; Visibility = 200 m; Visibility = 400 m

5 Conclusions

Aiming at the urgent needs of fog simulation technology, this paper proposes a fog simulation method based on improved self-supervised monocular depth estimation. Based on the atmospheric scattering model, detailed depth images were generated by multi-scale feature fusion and improved *manydepth* model. Then, the transmittance map was obtained by given visibility, and the dark channel map was used to estimate the atmospheric light value by distinguishing the sky region, so as to generate simulated fog datasets with different visibility. Through the comparison of 15 groups of experimental data, the fog simulation image conforms to human visual effect. This method is suitable for outdoor images with a small proportion of sky background, and its average fog simulation error rate is 6.2%, which indicates that the fog simulation method proposed in this paper is reliable and can solve the problems of lack of data sets and missing visibility data in fog days. It can be widely used in fog image recognition, visibility detection and fog target detection and other fields.

Acknowledgements. We acknowledge the support of Changsha University of Science and Technology Graduate Research Innovation Program (Grant No. 1208035).

References

1. Horvath, H.: On the applicability of the Koschmieder visibility formula. Atmos. Environ. **5**(3), 177–184 (1967)
2. Garg, R., Bg, V.K., Carneiro, G., et al.: Unsupervised CNN for single view depth estimation: geometry to the rescue. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. ECCV 2016. LNCS, vol. 9912, pp. 740–756. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_45
3. Godard, C., Mac Aodha, O., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 270–279 (2017)

4. Mansour, M., Davidson, P., Stepanov, O., et al.: Relative importance of binocular disparity and motion parallax for depth estimation: a computer vision approach. Remote Sens. **11**(17), 1990 (2019)
5. Chen, W., Fu, Z., Yang, D., et al.: Single-image depth perception in the wild. Adv. Neural Inf. Process. Syst. **29** (2016)
6. Nayar, S.K., Narasimhan, S.G.: Vision in bad weather. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 820–827 (1999)
7. Song, H.S., Liu, Y.P., Zheng, H.L., et al.: Road visibility detection based on priori theory of dark and bright primary colors. Laser Optoelectron. Progress **58**(6), 94–100 (2021)
8. Wang, J., Liu, Y.Y., Zhang, X.W., et al.: Summary of atmospheric light value estimation methods in haze images. Laser J. **42**(9), 6–10 (2021)
9. Yang, Y., Lu, X.X.: An image dehazing method combing adaptive brightness transformation inequality to estimate transmittance. J. XI'AN JIAOTONG Univ. **55**(6), 69–76 (2021)
10. Yu, M.J., Zhang, H.F.: Single-image dehazing based on dark channel and incident light assumption. J. Image Graph. **19**(12), 1812–1819 (2014)
11. He, K., Jian, S., Tang, X., et al.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
12. Watson, J., Mac Aodha, O., Prisacariu, V., et al.: The temporal opportunist: self-supervised multi-frame monocular depth. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1164–1174 (2021)
13. Long, X., Liu, L., Li, W., et al.: Multi-view depth estimation using epipolar spatio-temporal networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8258–8267 (2021)
14. Choi, S., Park, J., Yu, W.: Resolving scale ambiguity for monocular visual odometry. In: 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 604–608. IEEE (2013)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (2015). abs/1505.04597
16. Mishra, P., Herrmann, I.: GAN meets chemometrics: segmenting spectral images with pixel2pixel image translation with conditional generative adversarial networks. Chemom. Intell. Lab. Syst. **215**, 104362 (2021)
17. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. **21**(12), 4695 (2012)
18. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Process. Lett. **20**(3), 209–212 (2013)
19. Venkatanath, N., Praneeth, D., Chandrasekhar, B., et al.: Blind image quality evaluation using perception based features. In: Proceeding of the 21st National Conference on Communications, pp. 1–6, Washington, USA. IEEE (2015)
20. Yao, Y., Abidi, B., Doggaz, N., et al.: Evaluation of sharpness measures and search algorithms for the auto focusing of high-magnification images. Phys. A Stat. Mech. Appl. **6246**, 62460G-62460G-12 (2006)
21. Li, Y.F., Chen, N.N., Zhang, J.C.: Fast and high sensitivity focusing evaluation function. Appl. Res. Comput. **27**(4), 1534–1536 (2010)
22. Sen, A.: Quantum entropy function from AdS 2/CFT 1 correspondence. Int. J. Mod. Phys. A **24**(23), 4225–4244 (2009)
23. Song, S., Chandraker, M.: Robust scale estimation in real-time monocular SFM for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1566–1573 (2014)
24. Yuan, W., Gu, X., Dai, Z., et al. New crfs: Neural window fully-connected crfs for monocular depth estimation. arXiv preprint arXiv:2203.01502 (2022)

25. Mahmud, R., Buyya, R.: Modelling and simulation of fog and edge computing environments using iFogSim toolkit. Fog edge Comput. Princ. Paradig. 1–35 (2019)
26. Wang, Y.K., Fan, C.T.: Single image defogging by multiscale depth fusion. IEEE Trans. Image Process. **23**(11), 4826–4837 (2014)
27. Graffieti, G., Maltoni, D.: Artifact-free single image defogging. Atmosphere **12**(5), 577 (2021)