A SIFT-FREAK Based Framework for Coastline Image Stitching

Xiaoyu Wu¹,²*, Gong Chen², Xiangqian Ding¹, Kezhen Xie¹ and Guanqun Wei¹

¹Ocean University of China, Qingdao 266000, China;
²Shandong Computer Science Center, Jinan 250000, China
*Corresponding Author Email: wuxiaoyu628@126.com

Abstract. The quality of the images collected by the coastal zone video surveillance equipment is seriously degraded due to the sea fog, which directly affects the analysis of the image. Therefore, the study of the coastal image dehazing method is of great significance to the related research of the coastal zone. Coastal image has the characteristics of large sky area and monotonous color. The traditional method based on atmospheric scattering physics model is not suitable for this kind of image for block effect and color distortion. In this paper, we introduce the generative adversarial mechanism into sea fog image defogging, and propose a coastal image dehazing network based on it. The proposed model includes a generative network and a discriminative model, and is trained by adversarial mechanism. The generative model is composed of multi-scale feature extraction module and residual connection module. The discriminative network consists of two sub-networks of receptive field of different sizes.

1. Introduction

The vision system plays an important role in monitoring the scenes around the coastal zone. At present, the image acquisition of the coastal zone vision system generally depends on the visible light and infrared cameras. For the visible light camera, the scattering of atmospheric particles under sea fog will result in low contrast, color offset and distortion, decreased clarity, blurred texture and other image degradation phenomena in the image, which will seriously affect the accurate analysis of the image captured in coastal zone. Therefore, coastal zone image defogging is attracting more and more attention and is a promising research field.

Image defogging algorithms can be divided into physical model-based methods and learning-based methods [1]. The traditional hypothesis prior method estimates the image transmission rate through the artificial design hypothesis or prior information, and then restores the fog-free image based on the atmospheric scattering model. Such algorithms use optical-physical models and have strong pertinence. But the simplified physical model may lead to estimation error of global atmospheric light and image transmission rate. In recent years, deep learning, as the most popular direction in the field of artificial intelligence, has made great breakthroughs in image feature extraction, image recognition and other fields. It shows strong ability of learning and feature representation. Thus, some researchers have applied it to the image defogging task of land scenes [2-3].

However, most of the coastal zone images mainly include the sky region and the sea region, with monotonous colors. The traditional defogging algorithm based on dark channel prior has the problem of block effect and color distortion in the sky region. In this paper, a coastal zone image defogging model based on generative adversarial mechanism is proposed, and no physical model is adopted. The model includes a generative model and a discriminative model. The discriminative model determines whether the sample is from the data set or synthesized by the generative model. The generative model
makes the generated image indistinguishable as much as possible. The generative model, composed of multi-scale feature extraction module and residual connection module, is for feature extraction and fusion. It is used to estimate dehazed image. The discriminative model adopts two subnetworks of multiscale receptive field to discriminate the dehazed image.

2. Related Work

At present, the dehazing method mainly includes 2 categories. One is based on physical model, and the other one is based on image enhancement. The image degradation principle of fog is considered in the method based on physical model. It can restore the scene information with no fog by modeling and analyzing the atmospheric scattering. The dehazing method based on physical model mainly includes the algorithm based on the polarization characteristics of atmospheric light [4-5], the algorithm based on scene depth information [6-7], and the algorithm based on prior information [8-10]. The enhancement for hazy image does not study the principle of the effect on the image formation by haze. It improves the visual effect of degraded image by enhancing the contrast of image, enriching the details of image and correcting the color of image in image enhancement, instead of removing the influence of haze in the image. Nowadays, commonly used image enhancement methods include contrast enhancement algorithm [11-12] and Retinex theoretical enhancement algorithm [13].

Different dehazing algorithms have different adaptability and processing effects for hazy images of the same scene. The same image dehazing algorithm can achieve different results for different scenes. Most of the above defogging algorithms are proposed for land fog scenes, and scholars have studied some dehazing methods for sea fog scenes. Ju et al. [14] proposed a multi-scale cascade network to restore clear images from foggy sea surface images. Stojanović et al. [15] improved the atmospheric light estimation in atmospheric scattering model, making it more suitable for Marine monitoring image dehazing. Overall, there are few studies on sea fog image defogging.

Compared with the traditional defogging method based on physical model, the method based on neural networks can extract more abundant features and comprehensive information of images. The dehazing of coastal zone images will be better resolved by introducing neural networks.

3. The Proposed Network and Discussions

Inspired by [16], the proposed model adopts generative adversarial mechanism to dehaze the coastal zone images without any physical model. The establishment of the model is mainly divided into two stages: training stage and testing stage. It optimizes the model in training stage and outputs the dehazed image from the input hazy image in the testing phase.

The proposed model consists of a generative model and a discriminative model. The generative model is for multi-scale features extraction, fusion and dehazed image estimation by multi-scale feature extraction module and residual connected module. The discriminative model distinguishes the generated dehazed image by multi-scale receptive fields. The generative discriminative mechanism, generative model and discriminative model will be introduced below.

3.1. Generative Adversarial Networks Recap

Generative Adversarial Networks [17] is a deep learning model, consisting of generative model and discriminative model. The output is generated by game learning between the two models. The discriminative model determines the sample from dataset or generated by generative model. The generative model makes the generated image indistinguishable as much as possible. The two network models compete with each other to improve their algorithm capabilities until the discriminative model cannot tell whether the generated image is ground-truth image or not.

The generative model learns the distribution from a ground-truth image \( y \), it can be formulated as below:

$$
\min_{G} \max_{D} \mathbb{E}_{y \sim P_{data}(y)}[\log D(y)] + \mathbb{E}_{z \sim P_{z}(z)}[\log(1 - D(G(z)))] \tag{1}
$$

\( z \) is the noise input to generative model \( G \), \( G(z) \) represents the image generated by \( G \). \( D(y) \) is the probability that the discriminative model \( D \) judge \( G(z) \) to be a ground-truth image, and \( E \) is the
mathematical expectation. $y$ is ground-truth, and $\mathbb{E}(y)$ is expected to be 1. $\mathbb{E}(D(G(z)))$ is the probability that the discriminative model $D$ judge $D(G(z))$ to be a ground-truth image. $D$ wants $\mathbb{E}(D(G(z)))$ to be as large as possible, at the meantime, $V(D, G)$ will become small. $D$ hopes $\mathbb{E}(y)$ to be large and $\mathbb{E}(D(G(z)))$ to be small. Such a training task can be considered as a minimax game with a value function $V(D, G)$. When the discriminator cannot distinguish between the ground-truth image and the composite image, the training process is stopped to achieve a balance of the decision error between the generator and the discriminator.

![Architecture of the generative network](image)

**Figure 1.** Architecture of the generative network.

### 3.2. Generative Network

Generative Network generate a dehazed image from a single hazy coastal zone image without estimating the global atmospheric light and image transmission rate. As shown in Figure 1, the generative network includes an encoder and a decoder, consisting of multi-scale feature extraction module and residual connected module. The encoder decreases the size of the feature map. So the feature map of deep convolution layer has a smaller size and a larger receptive field, and contains more global feature information, such as the contour of the image. However, the feature map of shallow convolution layer has a larger size and smaller receptive field, and contains more local features.

Based on the above, the multi-scale features are fused to represent more rich information of the image. The proposed generative network changes the traditional way of transmitting image features at the same scale and fuses features of different scales, which is helpful for the network to adaptively select the features suitable for dehazing. The intersection in Figure 1 represents the multi-scale fusion process. The details are as follows: The encoder uses a convolution layer to the input image to obtain the shallow features and passes them to the deep feature layer of the decoder, and passes the deep features to the shallow feature layer of the encoder. So as to fuse the features in different scales.
The convolution filter is $3 \times 3$ and activation function is ReLU except those in multi-scale feature extraction module and residual connected module.

(1) Multi-scale feature extraction module

The convolution kernel with traditional single size can only extract the features of fixed scale, which is not conducive to recovering the detailed structure of the image. To solve this problem, multi-scale feature extraction module [18] is used to achieve adaptively haze feature extraction from a number of receptive fields at different scales. As shown in Figure 2, the multi-scale feature extraction module has 2 branches with the convolutional filter of $3 \times 3$ and $5 \times 5$ respectively. They share the convolution features in the same layer in different branch. The convolution in this module is dilated convolution, which can achieve haze feature extraction of greater receptive fields without increasing the number of parameters. In this paper, the expansion coefficients of two dilated convolutions on each branch are set as 1 and 2 respectively. The convolution filters of two sizes and dilated convolution of different expansion coefficients are combined to realize multi-scale features extraction of the image. At last, the features extracted in different branch are added and a $1 \times 1$ convolution layer is adopted for feature fusion. This convolution layer can not only reduce the dimension, but also improve the representation ability of the module. The activation function after this convolution layer is ReLU. The activation function in this module is LReLU and followed by the batch normalization, expect the last layer.

The module adopts the structure of 2 branches, with the output feature map interconnected of each branch. Too many branches and too large convolutional filters will lead to problems, such as the increase of network parameters and slow training. Thus, 2 branches structure is adopted in this paper to balance the network performance and the number of parameters.

(2) Residual connected module

In order to better integrate the image features extracted by different convolutional layers and realize the reuse of image features, the residual connected module [19] is also used to extract haze features.

The residual connected module adds up the haze features extracted from each layer with those from the upper layer, making full use of the haze features learned from different convolutional layers, realizing feature fusion with the upper layer, and extracting richer sea fog features. As shown in Figure 3, the first three convolutions of the residual module are dilated convolutions with expansion coefficients of 1, 2, and 3 respectively. The filter of the dilated convolution is $3 \times 3$. At last, a $1 \times 1$ convolution layer is adopted to get the output of residual connected module. The activation function in this module is ReLU, followed with batch normalization.

Figure 2. Illustration of the multi-scale feature extraction module
3.3. Discriminative Network

Discriminative network is to distinguish the haze free image and the dehazed image and train against the generative network to improve the visual quality of the dehazed image. The haze free image and dehazed image are taken as positive and negative samples, and the haze image is reference information. The discriminative network is trained supervision. The output of the discriminative network represents the probability that the input image is a positive sample, ranging from 0 to 1. The objective of the training is that if the input is a positive sample, the output value is close to 1; if it is a negative sample, the output value is close to 0.

The discriminative network is composed of two different subnetworks. The two sub-networks detect residual haze and dehazed artifacts in the dehazed image from a multi-scale perspective independently, so as to reduce the influence of residual haze on the discrimination results at different scales. The results of the two scales can be complementary, which is beneficial to improve the performance of the discriminator.

Two subnetworks with convolution filter of 3×3 and 5×5 respectively score the input image with different receptive fields, and take the average of the two scores as the final output of the discriminative network. Except for the different convolution filter, the two sub-networks are composed of six convolution layers. The activation function of the first five layers is the ReLU function, and the last layer is the Sigmoid function. The Batch Normalization is adopted after the activation function of all convolution layers, except the last one.

Compared with the traditional image defogging algorithm, the coastal image dehazing model based on generative adversarial mechanism has its own advantages. (1) The proposed method is not based on the atmospheric scattering model and can avoid the inaccuracy of global atmospheric light estimation in the traditional algorithm. (2) The discriminative model can effectively solve the problem of dehazed artifacts and residual fog in the dehazed image by means of adversarial training. (3) More valuable haze features are retained by multiple multi-scale feature extraction and fusion.

4. Conclusions

In this paper, we introduce generative adversarial mechanism into coastal zone image dehazing. By analyzing the mechanism, a coastal zone image dehazing model is proposed and discussed. The proposed model adopts multi-scale feature extraction module and residual connection module to generate the dehazed image, and use multi-scale discriminative networks for training. It can enhance the detail information of the coastal zone hazy image and improve the contrast, clarity and visibility of the image.

5. References

[1] Ren W, Liu S, Zhang H, et al. Single image dehazing via multi-scale convolutional neural networks[C]//European conference on computer vision. Springer, Cham, 2016: 154-169.
[2] Rashid H, Zafar N, Iqbal M J, et al. Single Image Dehazing using CNN[J]. Procedia Computer Science, 2019, 147: 124-130.
[3] Yang C, Liu Z, Liu S, et al. An End-to-End Pyramid Convolutional Neural Network for Dehazing[C]//Chinese Conference on Image and Graphics Technologies. Springer, Singapore, 2019: 41-50.
[4] Shwartz S, Namer E, Schechner Y Y. Blind haze separation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, USA: IEEE, 2006, 2: 1984-1991.

[5] Namer E, Schechner Y Y. Advanced visibility improvement based on polarization filtered images[C]. Polarization Science and Remote Sensing II, San Diego, USA: SPIE, 2005, 5888: 1-10.

[6] Kopf J, Neubert B, Chen B, et al. Deep photo: Model-based photograph enhancement and viewing[J]. ACM Transactions on Graphics, 2008, 27(5): 116:1-116:10.

[7] Narasimhan S G, Nayar S K. Interactive (de) weathering of an image using physical models[C]. IEEE International Conference on Computer Vision Workshop on Color and Photometric Methods in Computer Vision (CPMCV), New York, USA, 2003: 1-8.

[8] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior[C]. IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC: IEEE Computer Society, 2009: 1956-1963.

[9] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341-2353.

[10] Fu H, Wu B, Shao Y, et al. Scene-awareness based single image dehazing technique via automatic estimation of sky area[J]. IEEE Access, 2018, 7: 1829-1839.

[11] Kim J Y, Kim L S, Hwang S H. An advanced contrast enhancement using partially overlapped sub-block histogram equalization. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(4): 475-484.

[12] Seow M J, Asari V K. Ratio rule and homomorphic filter for enhancement of digital color image[J]. Neurocomputing: 13th European Symposium on Artificial Neural Networks, 2006, 69(7): 954-955.

[13] Jobson D J, Rahman Z, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 1997, 6(7): 965-976.

[14] Ju Q, Li C, Sang Q. Single Sea Surface Image Dehazing Using Multi-scale Cascaded Convolutional Neural Network[C]//2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI). IEEE, 2018: 1-5.

[15] Stojanović B, Miločević S, Stanković S. Improved dehazing techniques for maritime surveillance image enhancement[J]. Serbian Journal of Electrical Engineering, 2018, 15(1): 53-70.

[16] Liu Yuhang, Wu Shuai, Image Dehazing Algorithm Based on Multi-scale Fusion and Adversarial[J]. Laser & Optoelectronics Progress, 2020, 57(06): 179-188.

[17] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. arXiv preprint arXiv:1406.2661, 2019.

[18] Li J, Fang F, Mei K, et al. Multi-scale residual network for image super-resolution[C]//Proceedings of 2018 European Conference on Computer Vision, 2018: 527-542.

[19] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.