IOP Conf. Series: Journal of Physics: Conf. Series **1487** (2020) 012018 doi:10.1088/1742-6596/1487/1/012018

A Survey on Video Dehazing Using Deep Learning

Yue Feng[1, *]

[1] Department of Electrical Engineering, Columbia University, NY 10027, USA

[*] yf2466@columbia.edu

Abstract. With the fast development of object recognition and detection in autonomous driving and video monitoring, image with haze or raindrop can affect the result a lot. As deep learning develops, the hazed and raindrop image can lower the accuracy of object recognition and detection significantly. While hazed images cannot be managed using other image refining process. Since the noise in hazed image is signal-dependent. The object degradation in hazed image is related to object depth. So, the dehazing process depends on the input image. This paper provides a survey on single image and video dehazing methods, from end-to-end system to distributed system. General methods based on deep learning of state-of-art papers from 2010 to 2018 are summarized and compared, accompanied with their datasets of the current progress in this field. The application of these methods and relationship between these methos are also discussed in this paper.

1. Introduction

The development of video surveillance and processing allows researchers to enjoy the great application of artificial intelligence. Video dehazing system, as the effective tool to handle the problem of video processing in the application of robotics, has attracted the attention of researchers in this field. What makes dehazing image different from other image refining process is that the noise in hazed image is signal-dependent, which means that the degradation of objects in image depends on their depth in the image. Thus, the dehazing process also depends on input image.

The dehazing system is normally divided into two categories: image dehazing and video dehazing. In the dehazing process, an end-to-end system means that the input of the system is hazed image or video, and output is a dehazed image or a dehazed video. The dehazing model is generally used to acquire and represent the image data with haze and collect information through the explicit way and the implicit way. Moreover, the main dehazing model refers to the characteristics of the haze in images and frames. Different types of dehazing systems must consider the characteristics of the haze in images and frames, such as image-based dehazing system need to focus on at least colors, lights, and textures.

More and more recent researchers mainly concentrate on the dehazing algorithms to discover the atmospheric light's characteristic in image data and display the similar features of dehazing application. An end-to-end dehazing network designed by Li *et al.* [1] focuses on optimizing a reconstructed atmospheric scattering model which combines two parameters and input image into one feature *K*. This network uses a light-weight convolutional network which makes itself easy to implant to other models. Li *et al.* [2] also introduce an end-to-end video dehazing network that makes good use of the steadiness of video. Besides, they also trained the network together with object detection, and prove that the result is more accurate and robust. A benchmarking of single image dehazing by Li *et al.* [3] proposes several criteria to evaluate dehazing problems.

Depending on the above algorithms, many researchers combined the dehazing systems with the new deep learning algorithms in real applications. For example, Liu *et al.* [4] treat dehazing problem as an image restoration problem and propose a new loss function. They also introduce a domain-adaptive mask-RCNN to solve the object detection problem. Zhang and Patel [5] present a Multi-stream Dense Network (DID-MDN) that can classify the density of the raindrop and dehaze image according to classified rain-density. Their results are tested on two synthetic dataset and one real-world dataset. They also introduce an end-to-end Densely Connected Pyramid Dehazing Net-work (DCPDN) [6] on single image. This network learns parameters like transmission matrix and atmospheric light. A joint discriminator is proposed to monitor transmission map and dehazed image. And encoder-decoder structure is in introduced to learn transmission map. Qian *et al.* [7] use adversarial network to dehazing job. The generative network is trained to focus on raindrop area and the discriminative network are trained to evaluate local consistency of dehazed image. Ren *et al.* [8] use deep convolution neural network to do dehazing task on video frames by assuming a prior knowledge of global semantics and using this information to predict transmission map. Kim *et al.* [9] present a new cost function to do image and video dehazing job. This kind of cost function includes the effect of contrast and information loss. In video dehazing, they also remove flickering effect by making transmission value consistent. Zhang *et al.* [10] propose a new frame work to estimate optical flow and transmission map to dehaze on video and single image. Especially, they use Markov Random Field (MRF) to get the spatial context in their algorithm. Chen *et al.* [11]'s work can do dehazing job while minimizing artifacts. For constrain the exemplifying of artifact, they use Gradient Residual Minimization (GRM).

The remaining of the paper is organized as follows. Section 2 introduces several datasets in experiments. Section 3 reviews the recent dehazing algorithms based on deep learning. Section 4 describes the applications of dehazing systems. Section 5 presents the conclusions and some future works.

2. Datasets

2.1. NYU Depth V2 Dataset[12]

The NYU Depth V2 contains RGB and depth image taken by Microsoft kinect. Images of this dataset are from video sequences with 20 to 30 frames per second. Missing values of images are filled. Each component of images are labeled and numbered as chair1, chair2, chair3 and so on.

2.2. TUM RGB-D Dataset [13]

This dataset was built to evaluate SLAM (Simultaneous localization and mapping) system. It includes RGB, depth data captured by Microsoft kinect and also ground truth data of the Kinect sensor's path. The data comes from video with 680*680 resolution and 30 frames per second. They also provide an evaluation algorithm to assess the predicted camera path.

2.3. ILSVRC2015 VID dataset [14]

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset is built to do large-scale object recognition. There are two labels in the dataset: 1) object detection: binary flag indicating the existence of an object 2) object localization: bounding box indicating the size and location of an object. For object detection, it has 200 fully labeled categories with 1.2 million training images, 50 thousand validation images and 100 thousand test images. For object localization, it has 1000 categories.

2.4. RESIDE dataset[3]

The Realistic Single Image Dehazing (RESIDE) dataset uses clean images from NYU Depth V2 dataset and Middlebury stereo dataset to synthesis hazed image. When generating hazed image, different parameters in atmospheric scattering model are set. Overall it has 13,990 synthesized hazed images in training set. This contains 520 testing image that spans indoor to outdoor scenario.

2.5. Middlebury stereo dataset[15]

The Middlebury stereo dataset contains 9 images of a stereo scene of a certain image size and 2 disparity maps as ground truth. It has three types of image size: full, half and quarter. This dataset overall has 2 scenes and 6 image sizes.

Some examples from above-mentioned datasets are illustrated in Figure 1, and the overview description of datasets are shown in Table 1.

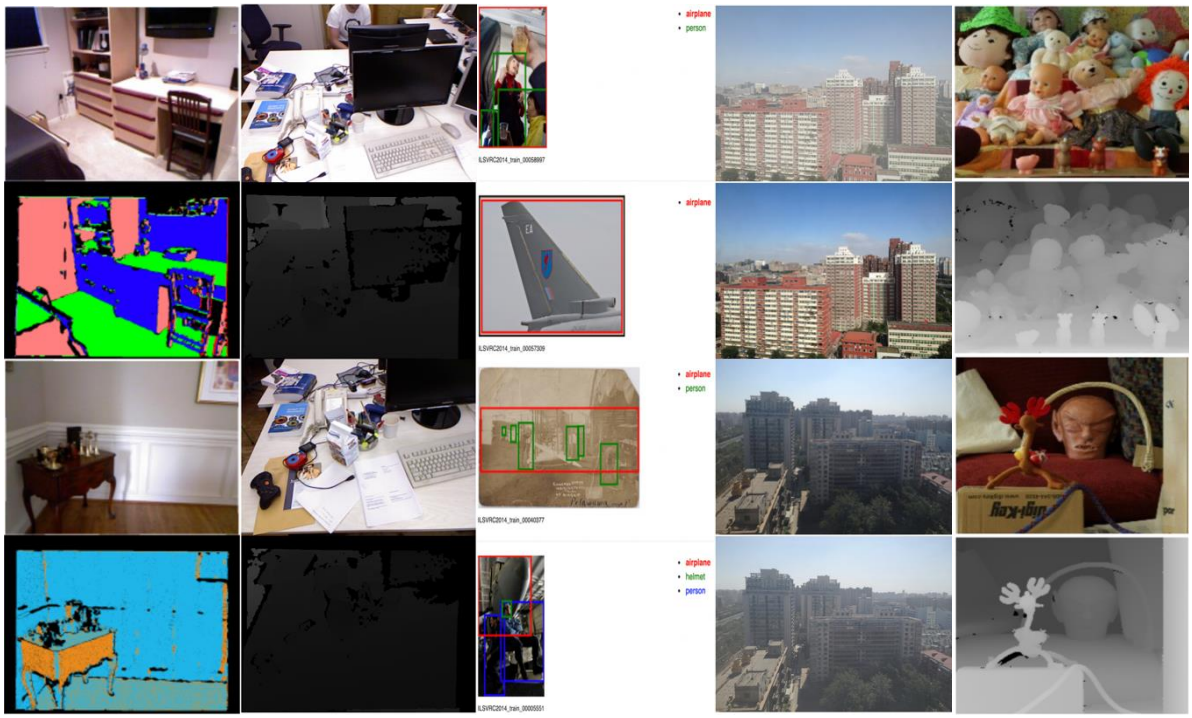

Figure 1. Example images from datasets, each row represents data from selected datasets (from left to right: NYU, TUM, ILSVRC, RESIDE, Middlebury)

Table 1. The description of five datasets for dehazing.

Dataset	# of images	Issues	Format
NYU Depth V2 Dataset	1, 449	video data with labeled objects and paired depth image	.mat
TUM RGB-D Dataset	/	color and depth images of Kinect sensor	PNG
ILSVRC2015 VID dataset	1, 500, 000	object location and label	JPEG
RESIDE dataset	86, 125	synthetic hazy images and clean image	JPEG/PNG
Middlebury stereo dataset	363	stereo image with ground truth disparity	PNG, pgm/ppm

2.6. Evaluation criterions

Six general quantitative criterions in the evaluation of dehazing performance are detailed as follows, such precision and recall (PR) curve, F-measure, and area under curve (AUC). To verify the effectiveness of the haze removal algorithm, the hazed images that contained bright reflective areas are always used to compare the processing performance. If the result of dehazing is good enough to

overcome the strong reflection light source, the visibility of dehazed images will be high and they are suitable for further image processing.

3. The Deep Learning Models

Recent researchers combined the dehazing systems with the new deep learning algorithms in dehazing system. Zhang and Patel[6]designed DCPDN to optimize the transmission map, atmospheric light and dehazing altogether. Atmospheric image degradation model is embedded to the optimized network. Encoder-decoder and multi-level pooling are introduced to deal with transmission map. This network uses a new loss called edge-preserving loss in optimization process. A joint-discriminator from Generative Adversarial Network (GAN) is used to estimate the structural correlation of transmission map and dehazed images.

The traditional method to describe hazing process is to build the atmospheric scattering model [16]. The hazed image depends on clean image, atmospheric light and transmission matrix. AOD-Net of single image dehazing proposed by Li et al. [1]combines atmospheric light and transmission matrix into one feature K by utilizing the information of clean image. It contains two part: 1) a K-estimation module which regress clean image from the hazed image 2) clean image generation module which generate dehazed image. The K-estimation gives the parameters to generate clean image. They use NYU depth dataset as ground truth to synthesize hazed image as training and testing dataset. EVDD-Net Model of video dehazing in Li et al. [2] get inspires from Li et al. [1]. It has 3 levels of fusion from input, K-estimation to output. This network is also integrated with a video object detection model. This is a light-weight network which converges very fast. The result shows that considering 5 consecutive frames can get the best result on object detection. They use both natural and synthetic video dataset.

To achieve higher precision and better performance, Li et al. [3]use PSNR/SSIM as full-reference parameters, spatial-spectral entropy-based quality (SSEQ) [17] blind image integrity notator using DCT statistics (BLIINDS-II) [18] as no-reference parameters to evaluate each state-of-art algorithm. They also use participants to evaluate algorithms by filling up a survey. This method evaluates algorithms from objective and subjective. Qian et al. [7] uses attentive GAN to remove raindrop. The generative network produce clean image which does not have raindrop. The discriminative network try to classify real or fake image. The generator uses LSTM and ResNet [19] to generate attention map. This discriminator uses several convolutional layers to get the result. DID-MDN network proposed by Zhang and Patel [5]contains two part: 1) residual-aware rain-density classifier which gives rain-density level of a certain image 2) multi-stream densely connected de-raining network which uses the information from 1) to generate images without rain streaks. After refinement, this network outputs a clean image. Liu et al. [4] proposed several loss functions in order to fulfill dehazing as degradation. And gives two solution sets to solve dehazing for detection. They try several combinations of the loss function and solution set and test the result.

To implement the image enhancement of dehazing performance in videos, Ren et al. [8] not only dehaze on video frames by assuming global semantic as prior, but also show that a stack of video frames can preserve the consistence without any assistance. They create a synthesis dataset based on NYU depth V2 dataset. The network they introduced uses a stack of 5 video frames to predict transmission maps of 3 video frames in the middle. Encoder-decoder structure is used in this network. Kim et al. [9] introduces an algorithm to do image and video dehazing by optimizing contrast. They first use quadtree-based subdivision to identify atmospheric light. And then predict transmission value to dehaze by adding information term in order not to saturated image and video. Zhang et al. [10]use human visualization system (HVS) and Markov Random Field (MRF) to obtain temporal and spatial consistency to do dehazing job. Their algorithm is also computational efficient by reducing interacting data and constrain minimum input data. Chen et al. [11]use local prior of haze image to predict transmission map. And use GRM to suppress the exaggeration of artifacts while recovering clean image. They also use Total Generalized Variation (TGV) [20] regularization to refine images. In these papers, deep learning methods are used to train the weights with a visible layer and a hidden layer that correspond to the input hazed image and the output dehazing image respectively.

4. Applications

Considering the recent developments of hot topics in dehazing systems, the network introduces by Ren et al. [8] can generate state-of-art result on video dehazing without tuning or image aligning. And the transmission map and be better maintained by their method in semantic segmentation. The algorithm designed by Kim et al. [9] can do dehazing on single image and video. This algorithm can avoid information loss while dehazing. AOD-Net proposed by Li et al. [1] are tested to help object recognition and detection on single image with tuning. Their group's EVD-net [2] on video dehazing can also boost object detection task and get higher average precision. Their result can be further used on autonomous driving and video monitoring scenario.

During the evaluation, Qian *et al.* [7] focus on raindrop dehazing with single image that can deal with images in bad condition. DID-MDN network proposed by Zhang and Patel [5] also aims at raindrop removal with rain density estimation. But there is no comparison with Qian *et al.* [7]'s work. DCPDN designed by Zhang and Patel [6] solves the problem of single image dehazing with different method in [5]. Liu *et al.* [4] try new loss function and test their network to improve object detection performance on single image with adaptive domain. Although these systems make dehazing successfully, the accuracy is similar to the deep learning based dehazing system, partly because a limited number of haze representation. Li *et al.* [3] also works on the same dataset of Liu *et al.* [4]. And they summarize several algorithms and state that further work can focus on no-reference metric developing. Zhang *et al.* [10] only improve the prediction of transmission map, and it behaves bad when image objects are similar to atmospheric light. Chen *et al.* [11] introduces a new image refinement algorithm to dehaze image while suppressing artifacts. Finally, these proposed approaches in dehazing applications outperforms the other state-of-the-art methods on the performance in terms of accuracy.

5. Conclusion

As the application of dehazing systems becomes more and more widespread, such as autonomous driving, image enhancement, and video monitoring. Video dehazing system has not only been an active area but also a challenging task. In recent years, a trend of the research is to address the issue by establishing deep neural networks and learning the latent features in order to model the information of atmospheric optical lights of raindrop and haze, which can seriously decrease the accuracy of object recognition and detection. In this paper, we present several dehazing systems based on deep learning techniques, which is devoted to extract the latent and explicit features of raindrops and hazes in images and videos. For raindrop removal, rain-density estimation can be applied to improve cleaning result. Similarly, for haze removal, some paper estimate transmission map and optical flow, while other uses GAN to do dehazing job. If the goal of dehazing is object detection and recognition, training the network combined with object detection and recognition is a good idea. As for evaluation, no-reference matrices are considered a prosperous direction. As the accuracy is improved, the cost-efficiency in practical dehazing application would be more important gradually.

References

[1] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). AOD-Net: All-in-One Dehazing Network. Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob, 4780–4788.
[2] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). End-to-End United Video Dehazing and Detection. (1), 7016–7023.
[3] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2019). Benchmarking Single-Image Dehazing and beyond. IEEE Transactions on Image Processing, 28(1), 492–505.
[4] Liu, Y., Zhao, G., Gong, B., Li, Y., Raj, R., Goel, N., … Tao, D. (2018). Improved Techniques for Learning to Dehaze and Beyond: A Collective Study.
[5] Zhang, H., & Patel, V. M. (2018). Density-Aware Single Image De-raining Using a Multi-stream Dense Network. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 695–704.

[6] Zhang, H., & Patel, V. M. (2018). Densely Connected Pyramid Dehazing Network. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3194–3203.
[7] Qian, R., Tan, R. T., Yang, W., Su, J., & Liu, J. (2018). Attentive Generative Adversarial etwork for Raindrop Removal from A Single Image. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2482–2491. https://doi.org/10.1109/CVPR.2018.00263
[8] Ren, W., Cao, X., Zhang, J., Xu, X., Meng, G., Ma, L., & Liu, W. (2019). Deep Video Dehazing with Semantic Segmentation. IEEE Transactions on Image Processing, 28(4), 1895–1908. https://doi.org/10.1109/TIP.2018.2876178
[9] Kim, J. H., Jang, W. D., Sim, J. Y., & Kim, C. S. (2013). Optimized contrast enhancement for real-time image and video dehazing. Journal of Visual Communication and Image Representation, 24(3), 410–425. https://doi.org/10.1016/j.jvcir.2013.02.004
[10] Zhang, J., Li, L., Zhang, Y., Yang, G., Cao, X., & Sun, J. (2011). Video dehazing with spatial and temporal coherence. Visual Computer, 27(6–8), 749–757. https://doi.org/10.1007/s00371-011-0569-8
[11] Chen, C., Do, M. N., & Wang, J. (2016). Robust image and video dehazing with visual artifact suppression via gradient residual minimization. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9906 LNCS, 576–591. https://doi.org/10.1007/978-3-319-46475-6_36
[12] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision, pages 746–760. Springer, 2012.
[13] Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; and Cre-mers, D. 2012. A benchmark for the evaluation of rgb-d slam systems. In Proc. of the International Conference on Intelligent Robot Systems.
[14] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115(3):211–252.
[15] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, vol. 1. IEEE, 2003, pp. I–I.
[16] E. J. McCartney. Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421 p., 1976.
[17] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," Signal Processing: Image Communication, vol. 29, no. 8, pp. 856–863, 2014.
[18] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," IEEE transactions on Image Processing, vol. 21, no. 8, pp. 3339–3352, 2012.
[19] K. He, X. Zhang, S. Ren, and J. Sun. Deep resid- ual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
[20] Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences 3(3), 492–526 (2010)