Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
14/1 (2021), 73-82. DOI: http://dx.doi.org/10.21609/jiki.v14i1.963

3D Information from Scattering Media Images

Laksmita Rahadianti

Faculty of Computer Science, Universitas Indonesia, Kampus UI Depok, 16424, Indonesia

Email: laksmita@cs.ui.ac.id

Abstract

Haze, fog, and bad weather conditions occur often in daily life. In these scattering media environments, micro-particles interfere with light propagation and image formation. Images captured in these conditions will suffer from low contrast and loss of intensity, hindering many computer vision methods. Thus, many approaches attempt to estimate the corresponding clear scene before processing the image further. However, the image formation model in scattering media shows that the 3D distance information is encoded implicitly in image intensities. In this paper, we provide a systematic review on methods to estimate relative depth and explicit depth directly from scattering media images. We use a dataset consisting of synthesized hazy images with known ground truths to establish accuracy, as well as real hazy images for a general visual analysis. For the accuracy evaluation, we demonstrate transmission estimation using statistical priors obtaining an average SSIM of 0.411 and MAE of 1.004; and depth map estimation using deep networks with an average SSIM of 0.305 and MAE of 0.860. Furthermore, for additional visual analysis, we also present the distance estimation for real hazy and underwater images of which we have no ground truth.

Keywords: *scattering media, 3D depth, transmission, statistical prior, airlight*

1. Introduction

3D distance estimation is a crucial step in computer vision applications to understand the structure of a captured scene. Many works have attempted to estimate the depth of a scene using either stereo pairs of images, monocular single images, or a combination of both [1]. While 3D distance estimation from stereo images has been studied extensively, the monocular approach is much more difficult.

In the case of single monocular images, most applications require geometric features from the image to aid the information extraction process [2]. In this effort, it is usually assumed that the image was captured in *clear media*, resulting in a clear image with distinct edges and objects, such as shown in Fig. 1a. However, in some cases, this assumption does not hold. In *scattering media* environments, the surrounding media will contain micro-particles that hinder light propagation. The interaction of light and micro-particles will cause *scattering* and *absorption* of light. As a result, the images captured by the camera in scattering media will not be good representation of the real scene. The captured images will present with low image quality, compromising the features necessary to extract information [3].

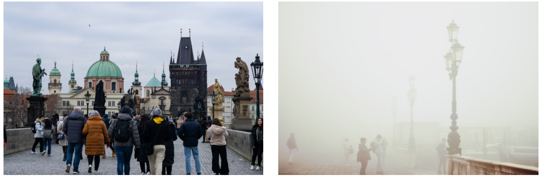

(a) Clear Image **(b)** Hazy Image

Figure 1. Examples of scattering media images

Images captured in environments such as fog, smoke, or light rain are often also referred to as *hazy images*. In these environments, the captured image suffer from low contrast, loss of detail, and the entire image displays a veiling effect with a white hue [4, 5]. Fig. 1b shows an example of images captured in these conditions. In scattering media environments, the appearance of the scene are severely compromised, making the features necessary for computer vision methods very difficult to find. To overcome this issue, most methods try to approximate a corresponding clear image before extracting its distance information [6, 7].

The physical process of scattering shows that the scattering effects actually contain valuable distance information. Thus, there is a significant amount of potential information that can be extracted directly from scattering effects, instead of treating them as noise. If so, the additional process of estimating a clear image can be bypassed completely. In this paper, we provide a systematic review comprising of various methods that can demonstrate the potential 3D distance information of relative depth and explicit depth maps from single hazy images. Specifically, we show the implementation of statistical priors to estimate relative depth, and deep networks to estimate explicit depth.

2. Images in Scattering Media

Scattering media environments are very different from their clear counterparts. In scattering media, there are micro-particles in the surrounding media which effect light propagation and as a result, image capture. Images captured in scattering media will thus show a very different appearance as well.

2.1. Image Appearance

In the hazy image in Fig. 1b we can observe the appearance of scattering media images. These images contain blurring effects, additive scattering noise, veiling effects, detail occlusion, loss of intensity, and low contrast.

Blurring effects occur when the image sensor at a certain pixel location receives a stimulus from more than one light ray [8]. In scattering media, the micro-particles in the environment will interfere with travelling light by altering their path [4, 5]. This is called *scattering*. These additional stray light rays are then captured by the camera causing blurring. The scattered light will then manifest in the image a veiling effect that obscures the scene. In hazy conditions, this veiling effect appears with a whitish hue. This veiling effect may be dominant in some local areas and the pixel intensity becomes saturated, occluding image details [9].

As explained in Section 1, the scattering media environment also causes *absorption*. The absorption reduces the intensity of the original image intensities from the scene that is able to arrive at the camera. Information on the color, edges, and shape of objects in scene are diminished by the time the light is captured. The loss of detail combined with the additive scattering noise represents an overall loss of intensity and contrast. This is significant as contrast is a basic perceptual attribute of an image carrying significant information [10].

2.2. Image Formation Model

The physical process of scattering and absorption is illustrated in Fig. 2. The image formation model used to approximate this process is based on the atmospheric scattering model [11]. In order to simplify the model, the scattering media is assumed to be homogeneous with a relatively low density [7].This assumption is reasonable for most natural settings in light weather conditions and non-murky natural bodies of water. The image captured in hazy conditions can be written as follows:

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

where, J is the original intensity, A is the *airlight* and t is the transmission of the surrounding media.

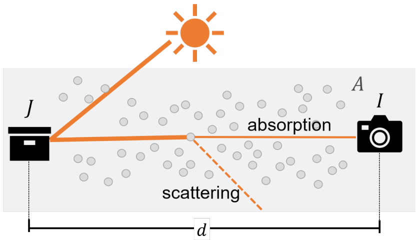


Figure 2. Image formation in scattering media.

The image formation model in Eq.(1) encapsulates both the scattering and absorption processes in scattering media environments. The left hand term describes the proportion of the original scene that successfully arrives at the camera after being subject to absorption, while the right hand term describes the additive scattering effects in the image.

2.3. Airlight and Transmission

The light propagation and image formation model in Section 2.2 introduces two new terms specific to scattering media images, namely *airlight* and *transmission*.

In scattering media, there is the aggregation of stray intensities from the light source and the scattering media. This represents is the color of the ambient light in the surrounding media, and is called airlight A [12]. The exact color of airlight in every scene differs based on the conditions and components of the surrounding media, as mentioned in Section 1. In natural hazy images, it is assumed that airlight is saturated at areas with the furthest distance $\approx \infty$ in natural outdoor scenes. These areas usually correspond with the areas of the background, or sky, in natural images. As an illustration, Fig. 3 shows an example of airlight areas in an scene, outlined in red.

The term *transmission* refers to the amount of light able to pass through the media, which is largely affected by the type of scattering media involved. In the event of image capture, the amount of any light intensity arriving at the camera will only be a fraction of the original intensity due to attenuation. Based on the Beer-Lambert law, transmission t can be written as follows:

$$t = e^{-\beta . d} \tag{2}$$

where β is the scattering coefficient of the media, and d is the distance to the camera.

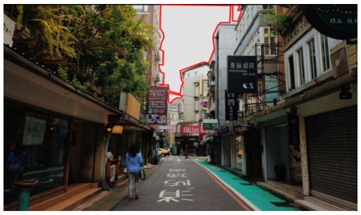


Figure 3. An example of a scene and its airlight areas outlined in red.

3. 3D Information from Scattering

The scattering effects in scattering media images are often treated as additive noise, and removed prior to any processing. However, based on Eq.(1), it is clear that the scattering effects themselves may carry important distance information. The captured image intensity is a direct function of 3D distance d, as shown in Eq.(2). This distance denotes the position of objects from camera, which is crucial for establishing the 3D structure of the scene. In this section, we will discuss the potential information that can potentially be extracted from scattering media images.

In the attempt to understand the 3D structure of the scene, it is ideal to obtain the exact values of distance. However, the exact value of explicit 3D depth d is very difficult to extract from a single hazy image I itself. For some computer vision applications, it may be sufficient to understand relative depth in the form of transmission t. We will review some potential methods to extract the relative depth in Section 3.1 and methods for extracting explicit depth in Section 3.2, such as shown in Fig. 4.

3.1. Relative Depth from Statistical Priors

The image formation model in scattering media presented in Section 2.2 show an inverse relation between intensity and depth. The image formation equation involves various unknown variables, so additional constraints are necessary. From Eq.(1) and Eq.(2) we attempt to exact this information using assumptions from statistical priors.

The **Dark Channel Prior (DCP)** was proposed based on observations of the general appearance of natural hazy images by He, et al. [13]. The work found that in patches of non-sky regions of natural hazy images, we can find *dark pixels* with very low intensity ≈ 0 in at least 1 image channel R, G, or B. This *dark channel* J_{DCP} is present in any clear image J as follows:

$$J_{DCP}(x) = \min_{s \in \{R,G,B\}} \Big(\min_{y \in \Omega(x)} \big(J^s(y) \big) \Big) \tag{3}$$

where x denotes the pixel location in the image, and $\Omega(x)$ denotes the local patch of pixels centered at x.

Based on the DCP principle, the dark channel will be very low ($J_{DCP} \approx 0$) only in **clear** natural images. However, in hazy images the dark channel will have higher intensities due to the additional scattered airlight. Thus, the the value of J_{DCP} will grow proportionally with the amount of additional scattering effects, which in turn increases with the distance from the camera. Thus, DCP can be used as an indicator of relative depth which can give a general 3D structure of the scene.

Conversely, Eq.(2) shows that transmission is inversely proportional to 3D depth, and hence can be used to indicate relative the position of objects. Relative depth is a useful cue for understanding the general structure of the scene. An example of a scene and its corresponding transmission map is shown in Fig. 5a and 5b.

It is possible to use DCP to estimate transmission t from a scattering media image [13]. This derivation assumes the airlight A is known, and that the t is constant in the local patch Ω. From Eq.(1), we divide each term with airlight A, apply a minimum operation on patch Ω and another minimum operation on each channel, arriving at:

$$\min_{s \in \{R,G,B\}} \Big(\min_{y \in \Omega(x)} \big(\frac{I^s}{A^s} \big) \Big) = \hat{t} . \min_{s} \Big(\min_{y \in \Omega(x)} \big(\frac{J^s}{A^s} \big) \Big) + (1 - \hat{t}) \tag{4}$$

Based on the DCP in Eq.(3), the first term on the right of is roughly equal to 0 (≈ 0), which allows us to estimate transmission based on DCP as follows:

$$\hat{t}_{DCP} = 1 - \min_{s \in \{R,G,B\}} \Big(\min_{y \in \Omega(x)} \big(\frac{I^s(y)}{A^s} \big) \Big) \tag{5}$$

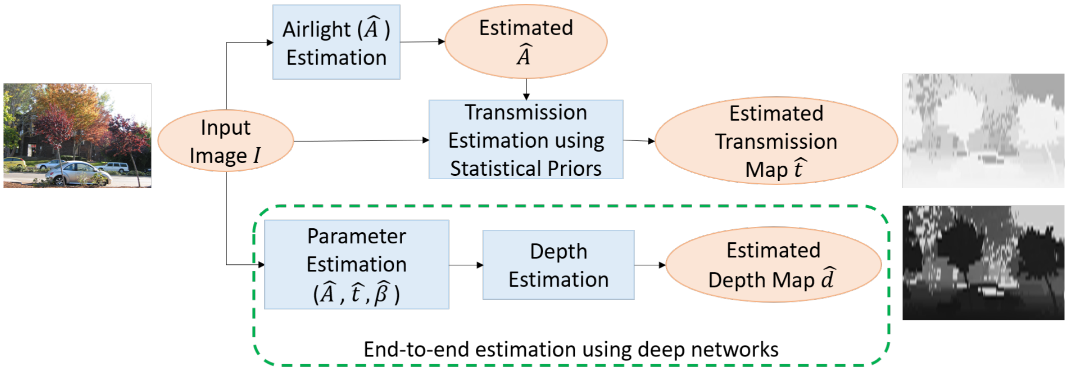

Airlight ($\widehat{A}$) Estimation
Estimated $\widehat{A}$
Input Image I
Transmission Estimation using Statistical Priors
Estimated Transmission Map $\widehat{t}$
Parameter Estimation ($\widehat{A}$, $\widehat{t}$, $\widehat{\beta}$)
Depth Estimation
Estimated Depth Map $\widehat{d}$
End-to-end estimation using deep networks

Figure 4. The flowchart of processes necessary to estimate relative depth (transmission) or explicit depth.

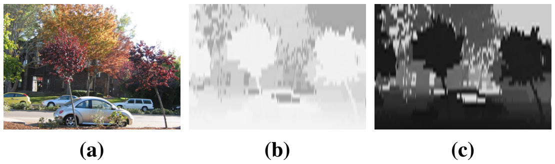


Figure 5. An example of a scene (a), transmission (b), and depth map (c).

Note that both DCP and transmission based on DCP will not give us values that represent the exact distance in units, but only a relative structure of depth. Assuming a wavelength-dependent scattering coefficient β, it is possible to estimate the direct depth d from the estimated transmission $\hat{t}$ in Eq.(1), given the scattering coefficient β is known. Unfortunately, the β coefficient is rarely known. Furthermore, β also varies greatly depending on the media, and the large variation of values makes this a poor general solution for scattering media images.

3.2. Depth Map Estimation using Deep Networks

In previous works, the transmission as described in Section 3.1 is used for *dehazing*, to recover the clear scene from the hazy image [12, 14]. The extraction of explicit 3D information in the form of a depth map is a much more complicated task. Due to the ambiguity, it is very difficult to map a single hazy image input to a corresponding depth map. An example of a scene and its corresponding depth map is shown in Fig. 5a and 5c.

To facilitate this level of image understanding from single images, it is necessary to extract higher level features [15]. In recent years, there have been many researches on deep learning networks for *dehazing* hazy images to their clear scenes [16–18]. There have been deep learning approaches that attempt to extract depth directly from images, but these researches assume clear images as inputs [19, 20]. We explicitly want to use hazy images directly of input, and obtain the 3D depth map that is clearly encoded in the intensities based on Eq.(1).

Image to image translation encodes the relationship between 2 groups, or domains, of images. The learned transformation can then be used to transform images from one domain to another. Deep learning architectures can be used to learn this transformation, enabling us to model 3D depth estimation as an image to image translation problem from hazy images to their corresponding depth map.

3.2.1. Conditional Generative Adversarial Networks.

The largest drawback of deep networks is the large amount of training data necessary to reach a stable state. Thus, a semi-supervised approach such as a Generative Adversarial Network (GAN) [21] is advantageous since it requires less data. A GAN framework involves both a generative and a discriminative network. The generator is trained to generate realistic reconstructions of the targets. Meanwhile, the discriminator is tasked to differentiate the generated reconstructions (fake images) from the ground truth (real images) [21]. Due to this adversarial process between the generator and discriminator, a GAN is capable of generating a final output which is a good visual reproduction of the target.

Pix2pix is a Conditional Generative Adversarial Network (cGAN) network for image to image translation, which can produce a distinct crisp output that reproduces the target well visually [22]. Pix2pix is particularly appropriate for image to image translation tasks. For general image to image translation tasks in computer vision, the high level goal is to obtain a *good* reconstruction that is indistinguishable from the reality [22]. This means that the resulting reconstruction might not have the minimum error, although it is visually similar. The general framework of the cGAN model used by Pix2pix is shown in Fig. 6.

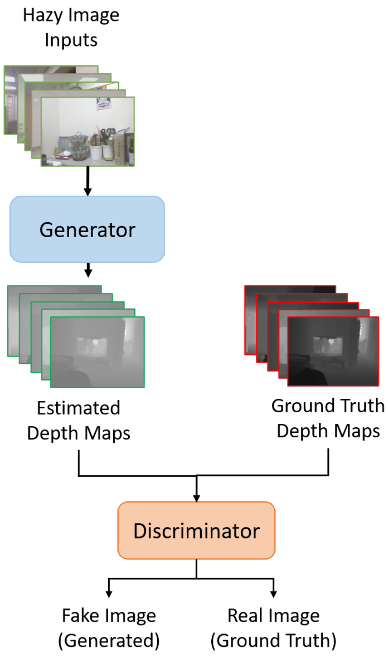


Figure 6. The cGAN framework of Pix2pix [22]

3.2.2. 2-Phase Depth Estimation of Hazy Images.

To obtain the best possible depth map estimation from hazy images, we use a 2-phase training approach [23]. Pix2pix is proven to be very powerful in image to image translation tasks, giving a good visual reconstruction of targets [22]. However, for image to depth tasks, accuracy is also important to estimate an explicit dense depth map of scenes. This error minimization may be better achieved using a more basic fully convolutional architecture for image to image translation, such as U-Net [24]. U-Net takes an input image into a series of convolutions to encode the image, followed by a series of transpose convolutions to decode the image back to the initial resolution. It also employs skip connections to bypass information directly from encoders to decoders on the same scale space. As with other fully convolutional models, U-Net is able to obtain an output with a minimized error.

The 2-phase training approach attempts to exploit both the U-Net and Pix2pix concepts to their advantage. In **Phase 1**, a U-Net architecture is trained as the generator within the cGAN framework of Pix2pix. Since it is semi-supervised, a smaller training set is sufficient at this stage of training. The final output of the trained network in Phase 1 is a visually similar depth map. Since this output is unlikely to have the lowest possible error, further training is necessary. In **Phase 2**, the saved pre-trained weights of the generative U-Net model from Phase 1 is used to initialize an independent U-Net architecture. This U-Net architecture is then trained independently to further minimize the reconstruction error of the output. At the end of Phase 2, the initial visually similar depth map will have been refined to minimize error. Thus, the final model should be able to reconstruct a estimated depth map that is both visually similar and accurate [23]. The 2-phase training approach is shown in Fig. 7.

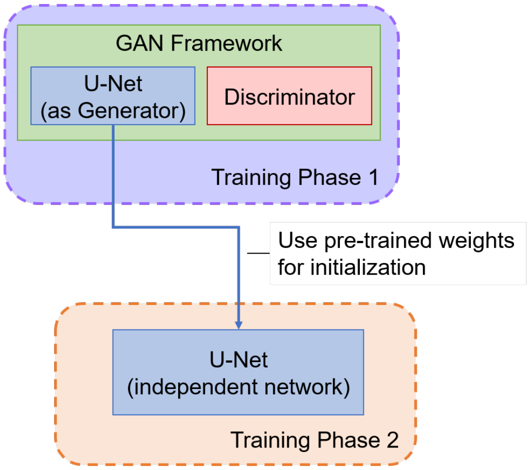


Figure 7. 2-Phase training for depth estimation [23]

3.3. The Special Case of Underwater Images

Underwater images are captured in underwater conditions such as through underwater photography or during search-and-rescue with underwater autonomous vehicles. The water surrounding the scenes in underwater environments also scatter and absorb lights, classifying those conditions as scattering media as well. Similar to their hazy counterparts, underwater images will also suffer from low image quality [6, 7]. It is necessary to distinguish underwater images compared to hazy images, due to the different hues of their scattering effects. While the scattering effects in hazy images usually are whitish, they appear as a blue-green hue in underwater images. This results in an additional effect of color distortion in underwater images.

As shown in Eq.(2), the transmission is a function of the scattering coefficient β. This coefficient is wavelength dependent. This spectral difference property is not very obvious in hazy images, thus it is often disregarded. Meanwhile, in underwater images the spectral differences are much more prominent, which is what causes blue-green veiling effect. The exact scattering properties of water varies depending on the mineral content, biodiversity, temperature, and other variables [25]. Thus, underwater images have a large variety of appearances and hues such as shown in Fig. 11a.

Due to the different appearances of underwater images, the airlight in an underwater image will also vary greatly. Note that the term *airlight* is used to refer to the ambient light for both hazy and underwater images. Due to this distinction, underwater images can not be treated similarly to their hazy counterparts and will be discussed separately in this paper.

It is not possible to use the same approaches for hazy images on their underwater counterparts. Drews, et al. [26] proposed the **Underwater Dark Channel Prior (UDCP)** is a modification of DCP for underwater images. Recall that underwater images are different than hazy images especially in their hue. Due to the wavelength dependency of the scattering coefficient in underwater scenes, underwater images will have diminished information in the R channel. While the dark pixel concept of DCP still holds, is only valid for the G and B channels, but not applicable to the R channel. For any clear image J, we can compute UDCP as follows:

$$J_{UDCP}(x) = \min_{s\in\{G,B\}} \Big(\min_{y\in\Omega(x)} \big(J^s(y)\big)\Big) \tag{6}$$

Galdran, et al. [27] proposed the **Red Channel Prior (RCP)** specifically to handle underwater images. The RCP handles the diminished R channel by taking its reciprocal channel $1 - R$ instead. Thus, in patches of non-water regions of natural underwater images, the *dark pixels* can be found in at least 1 image channel $1 - R$, G, or B. For any clear image J, we can compute RCP as follows:

$$J_{RCP}(x) = \min_{s\in\{1-R,G,B\}} \Big(\min_{y\in\Omega(x)} \big(J^s(y)\big)\Big) \tag{7}$$

Using the same derivation steps as in Eq.(4) and(5), it is possible to estimate transmission of underwater images using UDCP and RCP following Eq.(8) and (9).

$$\hat{t}_{UDCP} = 1 - \min_{s\in\{G,B\}} \Big(\min_{y\in\Omega(x)} \big(\frac{I^s(y)}{A^s}\big)\Big) \tag{8}$$

$$\hat{t}_{RCP} = 1 - \min_{s\in\{(1-R)G,B\}} \Big(\min_{y\in\Omega(x)} \big(\frac{I^s(y)}{A^s}\big)\Big) \tag{9}$$

4. Experiments

This section will demonstrate the extraction of potential information from scattering media images as described in Section 3. The information that will extracted and demonstrated here will start from the airlight, transmission (Section 3.1) and explicit depth maps (Section 3.2). Lastly, we will present some additional experiments that attempt to handle underwater images despite the wavelength dependency as discussed in Section 3.3.

4.1. Dataset

To conduct our experiments, we use 2 types of hazy image data, e.g. simulated hazy images and real hazy images. Real natural hazy images do not usually come with the corresponding ground truth information of depth. The creation of a real hazy image and depth dataset is largely attributed to the physical limitations of capturing depth data in hazy conditions. Some well-known tools to record depth have limitations of maximum depth and hardware constraints in environments in which microparticles are present. Thus, we use a simulated hazy dataset and a real hazy dataset to demonstrate the methods described in Section 3.1 and 3.2.

4.1.1. Simulated Hazy Dataset. To demonstrate accuracy and reliability of the extracted information as discussed in Section 3, we need to evaluate the results based on ground truth distance information. Thus, we first use a simulated dataset of hazy images from which the ground truth distance is available. We utilize the NYU depth dataset [28] which provides a dataset of real images and their corresponding pixelwise depth maps. Due to equipment limitations, the NYU depth dataset consists of images and depth maps captured indoors, as shown in Fig. 8a and 8b. Then, using the image formation model in scattering media in Eq.(1), we generate simulated hazy images such as that shown in Fig. 8c.

From the full set of of NYU images, we manually select scenes that have a variety of depths instead of a constant depth throughout the image. Since the NYU dataset consists of indoor images, we also remove scenes that include windows and/or additional light sources, in order to adhere to the assumption of a single light source in Eq.(1). The simulated images are generated with various values of airlight $A \in [0.7, 1]$ and scattering coefficient

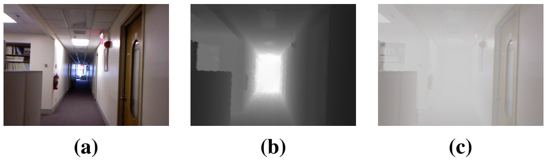


Figure 8. Original image (a), depth map (b), and Simulated hazy image (c) from the NYU Dataset [28].

$\beta \in \{0.1, 0.2, 0.3, 0.4\}$ [23, 29]. We obtain 700 simulated hazy scenes such as shown in Fig. 8c.

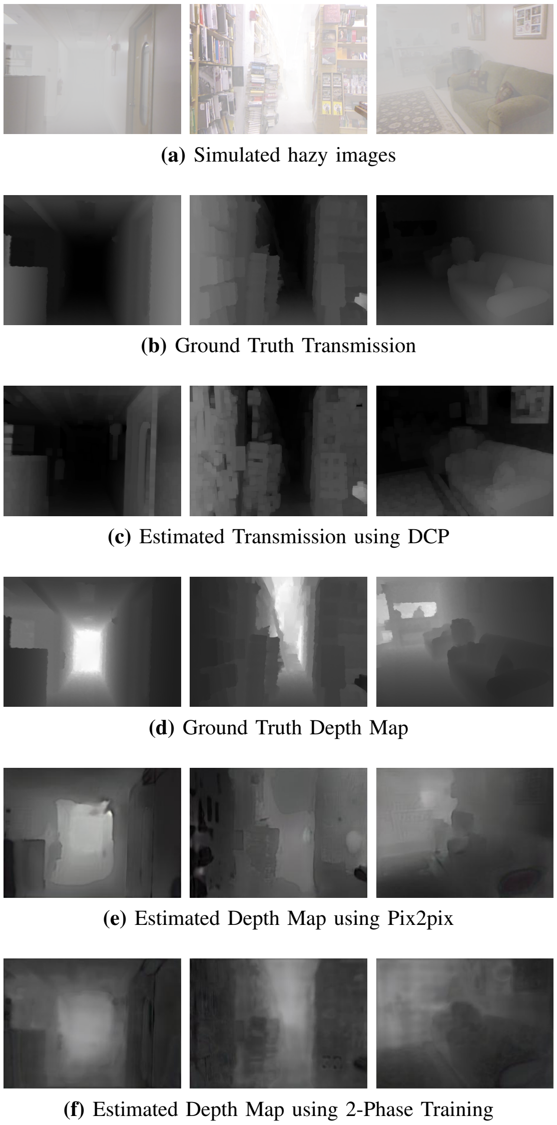
4.1.2. Real Hazy Dataset. To further analyze the methods discussed in Section 3, we then attempt to extract depth information from real hazy images. In this part we use the O-Haze dataset [30], which consists of 45 different hazy outdoor scenes and their corresponding haze-free images. The hazy images are captured with generated haze from 2 professional fog machines during overcast and non-windy conditions. Samples of images from the O-Haze dataset [30] are shown in Fig. 10a. The distance information of the scene is not available in this dataset.

4.2. Transmission Estimation

In this section, we will demonstrate the estimation of relative depth in the form of transmission. In the first part, we use DCP as described in Section 3.1 and the image to image translation concepts introduced in Section 3.2 on the simulated hazy dataset. The simulated dataset is used here because the ground truth transmission is available, enabling us to compute accuracy. The detailed process of transmission estimation is depicted in the upper section of the flowchart shown in Fig. 4.

Eq.(5) requires the airlight A to compute transmission. In the event that the airlight is unknown, there are some methods that can be used. Recall that J_{DCP} as an indicator of depth, while airlight should be found at the farthest location from the camera. Thus, airlight can be found at the location with the maximum value of J_{DCP}. However, in scenes where there are bright objects near to the camera, the DCP alone is insufficient. Thus, additional cues and computation may be necessary [12, 31]. In these experiments, we assume that the airlight A is known, so the transmission can be estimated using DCP based on Eq.(5).

To evaulate the accuracy of estimation we use three metrics, i.e the an mean absolute error (MAE) and the structural similarity (SSIM). The MAE is the pixelwise depth estimation error over the scene expressed in meters. A lower value of MAE shows a better estimate. The SSIM is a perceptual image quality metric based on the human visual system [32]. SSIM values range from -1 to 1, with a higher value indicating better visual quality.


(a) Simulated hazy images

(b) Ground Truth Transmission

(c) Estimated Transmission using DCP

(d) Ground Truth Depth Map

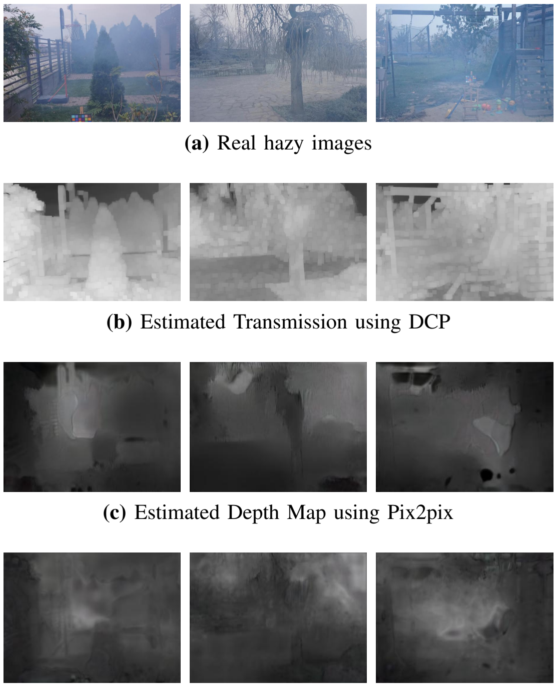
(e) Estimated Depth Map using Pix2pix

(f) Estimated Depth Map using 2-Phase Training

Figure 9. Estimated transmission and depth maps of simulated hazy scenes.

The transmission estimation is evaluated over the test set of 50 images, with a SSIM of 0.411 ± 0.101 and MAE of 1.004 ± 0.520. The results show that DCP is not able to give an accurate estimate compared to deep learning models. However, recall that transmission is a *relative* depth cue. Thus, we do not concern ourselves too much on the accuracy, as long as it is able to give us a general 3D structure of the scene. This structure is visible in the estimated transmission maps shown in Fig. 9c.

We also take the real hazy scenes from the O-Haze dataset [30] and estimate transmission based on DCP. The results are shown in Fig. 10b. These real images do not have a known ground truth transmission, thus we cannot compute accuracy. However, from a visual inspection, we can see that the estimated transmission can give us a good general idea about the 3D structure of the scene.



(a) Real hazy images

(b) Estimated Transmission using DCP

(c) Estimated Depth Map using Pix2pix

(d) Estimated Depth Map using 2-Phase Training

Figure 10. Estimated transmission and depth maps of real hazy scenes.

4.3. Depth Map Estimation

In this section, we move on to estimate the dense depth map of single hazy images. The depth map represents an explicit 3D distance estimate of the scene. As discussed, this ill-posed problem will be solved using a deep learning approach. We compare the performance of 3D distance estimation using the 2-phase training model described in Section 3.2.2 and Pix2pix. Once again, the simulated dataset is used because of the availability of the corresponding ground truth depth maps.

The 2-phase training network is trained with a training set of 650 hazy image - depth pairs, with the same number of image pairs as for estimating transmission, e.g. 200 hazy-depth pairs for phase 1 and 450 pairs for phase 2. Once again we compare the results of the 2-phase training approach with the Pix2pix GAN framework alone. The Pix2pix GAN is trained with all 650 training images in one go. Both networks are trained using SGD with a learning rate (LR) of 2×10^{-4} for 200 epochs. After training, the depth estimation is conducted on a test set of 50 images. The evaluation of depth estimation is shown in Table 1. For visual inspection, we also present the estimated depth maps in Fig. 9e and 9f.

Table 1. Evaluation of Estimated Depth

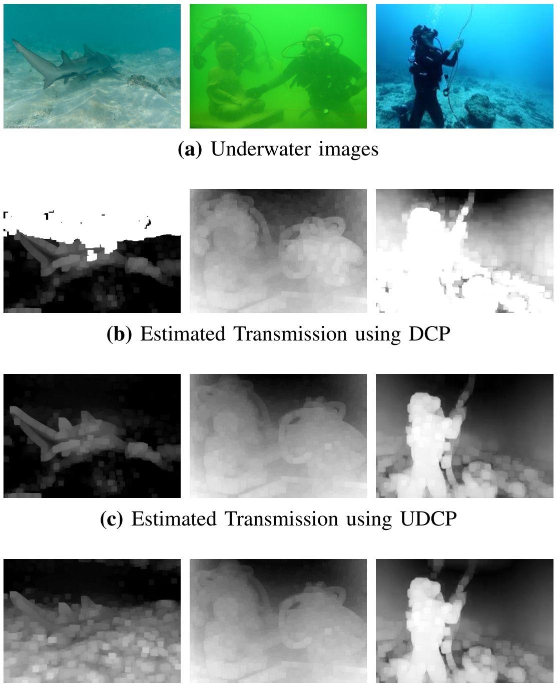
Method	SSIM	MAE
$\hat{d}$ Pix2pix [22]	0.321 ± 0.056	0.880 ± 0.339
$\hat{d}$ 2-phase [23]	0.305 ± 0.057	0.860 ± 0.464

Finally, to visualize the depth estimation of real images, we take hazy scenes from the O-Haze dataset [30] and estimate their depth maps. Although we cannot compute accuracy, we can see that the estimated depth is visually distinct and gives us the 3D structure of the scene. The detailed estimation results are shown in Fig. 10c and 10d. The results show that the trained model is adapted to the training set used to train it, and is poorly generalized for other hazy images. However, considering the previous results in Fig. 9, the deep image to image models show the capacity of estimating depth if provided with sufficient training data.

4.4. Challenging Underwater Images

In this section, we attempt to process underwater images and show the possible 3D distance information that can be obtained from them. We focus on the relative depth only using the statistical priors described in Section 3.3. We use the UDCP and RCP to estimate transmission of the underwater scene, and compare them to the DCP.The results displayed in Fig. 11 show that the DCP in Fig. 11b is incapable of handling the color scheme of underwater images. Meanwhile, as shown in Fig. 11c UDCP gives a better estimate, and RCP (Fig. 11d) gives an even better relative depth estimate of the scene.

In this paper, we do not attempt to extract the explicit depth map of underwater scenes due to the unavailability of data. In order to train the depth estimation models in Section 3.2, a good dataset of underwater images and their corresponding ground truth depths is crucial.



(a) Underwater images

(b) Estimated Transmission using DCP

(c) Estimated Transmission using UDCP

(d) Estimated Transmission using RCP

Figure 11. Estimated transmission $\hat{t}$ of underwater images.

5. Conclusion and Future Work

In this paper, we have demonstrated the potential distance information that can be extracted from a single image captured in scattering media. Other works usually perform *dehazing*, i.e. eliminating the scattering effects to obtain a corresponding clear image, from which further computer vision methods can be done to further extract information of the scene. However, based on the image formation model in scattering media environments, we propose the idea that the scattering effects themselves contain distance information of the scene. This paper explores the possibility of exploiting the scattering effects instead of removing them, and thus eliminating the dehazing process entirely.

Extracting distance information directly from single images is very difficult due to various unknown variables. In some applications, a relative depth is sufficient for a rough understanding of the objects and their positions relative to the camera. This paper implements various methods to extract relative and explicit depth, e.g. the Dark Channel Prior for estimating transmission as relative depth with an SSIM of 0.411 ± 0.101 and MAE of 1.004 ± 0.520, Pix2pix for explicit depth with an SSIM of 0.321 ± 0.056 and MAE of 0.880 ± 0.339, and the 2 phase training framework also for explicit depth with an SSIM of 0.305 ± 0.057 and MAE of 0.860 ± 0.464. These quantitative measures can be computed on the synthetic image dataset which includes known ground truth depths. We further attempt to estimate depth using real hazy images, but since they do not come with ground truth depth, we are only able to provide a subjective analysis of the images, which show a good visual estimate about the 3D structure of the scene.

Although we were able to provide a systematic review on the potential methods of 3D information extraction in this paper, there are still a lot of potential areas to build upon. The statistical priors used have been studied to create generalizations and improvements. Thus the estimation of individual parameters such as airlight, can also be continuously improved with the end goal of obtaining an even better depth estimate.

In our experiments, we find that the lack of data itself is a large issue. A realistic dataset with known ground truth 3D information is quite difficult to obtain. We have shown that the deep learning models show promising results in estimating depth maps of single hazy images through image to depth translation architectures. However, these deep models require a large amount of training data to reach stability, so a standardized dataset of image-depth pairs is necessary to enable reliable direct depth estimation using deep learning.

Finally, as an additional experiment, we attempt to challenge underwater images in this paper. Underwater images need to be treated differently from their hazy counterparts due to the complexity of the color transformation. We give a visual analysis on the applications of the Underwater Dark Channel Prior and the Red Channel Prior to estimate transmission specifically for underwater images. Previous research on hazy images has been quite extensive, however that is not the case for underwater images. Thus, there is a large potential for future work to be extended to underwater images.

Acknowledgement

This research was supported by Universitas Indonesia through Hibah Publikasi Terindeks Internasional (PUTI) Q3 No. NKB-1822/UN2.RST/HKP.05.00/2020.

References

[1] A. Saxena, J. Schulte, and A. Y. Ng, "Depth estimation using monocular and stereo cues," in *International Joint Conference on Artifical intelligence*, vol. 7, Jan. 2007, pp. 2197–2203.

[2] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5162–5170.

[3] S. G. Narasimhan and S. K. Nayar, "Interactive deweathering of an image using physical models," in *IEEE Workshop on Color and Photometric Methods in Computer Vision*, Oct. 2003.

[4] C. Tsiotsios, M. E. Angelopoulou, T.-K. Kim, and A. J. Davison, "Backscatter compensated photometric stereo with 3 sources," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 2259–2266.

[5] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, Jul. 2002.

[6] R. Schettini and S. Corchs, "Underwater image processing: state of the art of restoration and image enhancement methods," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, pp. 1–14, Apr. 2010.

[7] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 81–88.

[8] R. Wang and D. Tao, "Recent progress in image deblurring," *arXiv preprint arXiv:1409.6838*, 2014.

[9] Y. Li, H. Lu, K.-C. Li, H. Kim, and S. Serikawa, "Non-uniform de-scattering and de-blurring of underwater images," *Mobile Networks and Applications*, vol. 23, no. 2, pp. 352–362, 2018.

[10] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[11] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, 2008.

[12] C. O. Ancuti, C. Ancuti, and C. De Vleeschouwer, "Effective local airlight estimation for image dehazing," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 2850–2854.

[13] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, Sep. 2011.

[14] H.-H. Chang, C.-Y. Cheng, and C.-C. Sung, "Single underwater image restoration based on depth estimation and transmission compensation," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 4, pp. 1130–1149, 2018.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015, pp. 1–9.

[16] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multiscale convolutional neural networks," in *European conference on computer vision*. Springer, 2016, pp. 154–169.

[17] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7314–7323.

[18] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "Pdr-net: Perception-inspired single image dehazing network with refinement," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 704–716, 2019.

[19] B. Liu, S. Gould, and D. Koller, "Single image depth estimation from predicted semantic labels," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 1253–1260.

[20] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. of the 27th International Conference on Neural Information Processing Systems (NIPS)*, 2014, pp. 2366–2374.

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017.

[23] L. Rahadianti, F. Sakaue, and J. Sato, "Depth estimation from single hazy images with 2-phase training," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2020, pp. 309–316.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Oct. 2015, pp. 234–241.

[25] A. Morel, B. Gentili, H. Claustre, M. Babin, A. Bricaud, J. Ras, and F. Tieche, "Optical properties of the "clearest" natural waters," *Limnology and Oceanography*, vol. 52, no. 1, pp. 217–229, Jan. 2007.

[26] P. Drews, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *IEEE International Conference on Computer Vision Workshops*, Dec. 2013, pp. 825–830.

[27] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, Jan. 2015.

[28] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European Conference on Computer Vision*, Oct. 2012, pp. 746–760.

[29] C. Li, J. Guo, F. Porikli, H. Fu, and Y. Pang, "A cascaded convolutional neural network for single image dehazing," *IEEE Access*, vol. 6, pp. 24 877–24 887, 2018.

[30] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 754–762.

[31] L. Rahadianti, F. Sakaue, and J. Sato, "Time-to-contact in scattering media environments based on statistical priors," *ITE Transactions on Media Technology and Applications*, vol. 5, no. 4, pp. 147–161, 2017.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.