

Article

Novel Descattering Approach for Stereo Vision in Dense Suspended Scatterer Environments

Chanh D. Tr. Nguyen ¹, Jihyuk Park ¹, Kyeong-Yong Cho ², Kyung-Soo Kim ^{1,*} and Soohyun Kim ^{1,*}

¹ Department of Mechanical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea; chanhnguyen@kaist.ac.kr (C.D.T.N.); jihpark@kaist.ac.kr (J.P.)

² Unmanned Safety Robot Research Center, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea; yong00@kaist.ac.kr

* Correspondence: kyungsookim@kaist.ac.kr (K.-S.K.); soohyun@kaist.ac.kr (S.K.); Tel.: +82-42-350-3228 (K.-S.K. & S.K.)

Received: 19 May 2017; Accepted: 14 June 2017; Published: 17 June 2017

Abstract: In this paper, we propose a model-based scattering removal method for stereo vision for robot manipulation in indoor scattering media where the commonly used ranging sensors are unable to work. Stereo vision is an inherently ill-posed and challenging problem. It is even more difficult in the case of images of dense fog or dense steam scenes illuminated by active light sources. Images taken in such environments suffer attenuation of object radiance and scattering of the active light sources. To solve this problem, we first derive the imaging model for images taken in a dense scattering medium with a single active illumination close to the cameras. Based on this physical model, the non-uniform backscattering signal is efficiently removed. The descattered images are then utilized as the input images of stereo vision. The performance of the method is evaluated based on the quality of the depth map from stereo vision. We also demonstrate the effectiveness of the proposed method by carrying out the real robot manipulation task.

Keywords: low visibility; backscattering; descattering; defogging; stereo vision

1. Introduction

High spatial resolution ranging is crucial in robot manipulation and a depth map is necessary to accomplish the task. There are many cases where the system works in low visibility and strong scattering environments, such as underwater robots or firefighting robots. Our application is bipedal and quadrupedal robots working in nuclear power plants where they must cope with poor visibility due to dense steam. When an accident occurs, the plant is filled with very dense water-based atmospheric particles, and the robot needs to operate the plant. From our experiments, the commonly used sensors such as LiDAR (LMS511, SICK, Waldkirch, Germany and UTM-30LX-EW, Hokuyo, Osaka, Japan) and time of flight (ToF) camera (Kinect v2, Microsoft, Redmond, WA, USA) are unable to work in such low visibility conditions. Our conclusion is consistent with the study by Starr and Lattimer [1]. Some specialized subsea LiDAR systems (please refer to Massot-Campos and Oliver-Codina [2] for a comprehensive survey of underwater 3D reconstruction), laser line scanning [3] or structured light [4–7] are able to operate in scattering media. However, these systems are power-consuming, slow, and bulky and thus they are not well suited for a walking robot. Our goal is to utilize the images from a standard stereo vision system for robot manipulation in a scattering environment. Therefore, no additional hardware is required for the stereo vision system.

Stereo vision has been intensively studied for decades since retrieving the depth map of a scene is critical in many applications such as driving assistance and automated robotics. However, most state-of-the-art methods of stereo vision primarily deal with high-quality images from datasets, for

example, Middlebury datasets [8,9], and focus on either reducing the matching error or providing a real-time system [8,10,11]. Most stereo vision algorithms follow the multi-stage framework codified by Scharstein and Szeliski [8]. The rectified images should pass four main sequential steps to obtain the disparity map. The four stages are matching cost computation, cost aggregation, disparity selection, and disparity refinement. In general, the stereo algorithms can be classified into two categories, namely, local and global approaches. In a local approach, the disparity computation at given point depends only on the intensity value within the local window of grayscale images [12] or color images [13]. It has low computational complexity and short running time. These methods commonly have an inherent conceptual problem that it is assumed that the region inside the window is a fronto-parallel surface and does not cover the depth discontinuities. Studies based on varying support-weights window [14] or the geodesic support weights [15] can overcome the problem of depth discontinuities but are time-consuming. A more recent approach [16] that utilized guided filtering [17] achieved start-of-the-art results very efficiently. In a global approach, the problem, on the other hand, is formulated as a global optimization problem. In this approach, second and third steps are combined, and the main difference lies in how the optimization problem is solved. The problem can be solved efficiently using graph cut [18,19] and loopy belief propagation [20,21], among others. However, these methods, in practice, are rather slow. Both local and global methods use the photo-consistency constraint to find the corresponding pixels. In other words, they try to find the most similar pixel intensities in the left and right image.

These methods, however, cannot be applied directly to images taken in an indoor dense scattering environment where active light sources are required for illumination. The reason is that the scene radiance is attenuated while it propagates before reaching the camera. The greater the distance is, the weaker the signal that the camera receives is. Therefore, the image contrast is low. Additionally, the cameras capture the scattering signal, which increases with the object distance, scattered by the suspended particles. Furthermore, using non-parallel and non-uniform artificial illumination sources, situated close to the cameras, generates a significant backscattering signal, which is spatially-varying, under a geometric constraint. Thus, the intensities of the same object captured by two cameras of the system can be significantly different. Therefore, the photo-consistency does not hold. For the close range measurement in our case, stereo vision wrong matching is mainly due to backscattering rather than low contrast. Thus, non-uniformity of the backscattering is the dominant cause of wrong matching in stereo vision.

The stereo vision of a natural foggy scene can take advantage of developed image visibility enhancement methods. The polarization-based method enhanced the haze images under natural light [22] or underwater images utilizing active polarized light [23,24] by examining the degree of polarization (DOP) from multiple images taken under different polarization states. The methods are based on the assumption that the DOP of the object is spatial-invariant, which does not hold in our case. There has been significant progress in single image removal of haze, a process called dehazing, based on Koschmieder's law [25]. Markov random field (MRF) was used as a framework to derive the cost function by Tan [26], Fattal [27], and Nishino et al. [28]. Based on natural image statistics, the well-known Dark Channel Prior (DCP) was derived by He et al. [29]. Owing to DCP's effectiveness in dehazing, the majority of start-of-the-art dehazing techniques [30–32] have adopted the prior. Recently, learning-based methods [33,34] have also been utilized to solve image dehazing problems, providing state-of-the-art results. These methods, except for [23,24], targeted corrupted images primarily caused by attenuation rather than non-uniform backscattering from active illumination.

Recently, several nighttime dehazing algorithms have been developed. Zhang et al. [35] utilize a new imaging model to compensate light and correct color before applying DCP. Li et al. [36] incorporate a glow term into standard nighttime haze model. After the glow is decomposed from the image, DCP is employed to obtain a haze-free image. These methods can be utilized in a pre-processing step for stereo vision in a scattering scene with active light sources. However, they are not real-time capable.

Several methods were introduced to solve stereo vision for images of fog or underwater scenes. Caraffa and Tarel [37] combine photo-consistency term and atmosphere veil depth cues to formulate the problem and solve stereo and defog by utilizing the α -expansion algorithm [18]. This method is sensitive to the nonlinear camera response function and image noise. Therefore, the authors demonstrated proper results for synthetic images but not real foggy images. Roser et al. [38] iterate applying a conventional stereo algorithm to compute the depth and using depth to recover the object radiance. The method, however, does not model light scattering in the stereo matching step and defogs video frames independently, which cause errors in stereo matching. Li et al. [39] solve depth reconstruction and defog simultaneously from monocular video based on structure-from-motion (SfM). This only works when SfM can be calculated. Furthermore, the method is far from real-time capable since 10 min per frame is reported. These studies noted above are capable of processing images obtained under natural light sources only. Negahdaripour and Sarafranz in [40] use both photo-consistency and backscattering cues to estimate disparity by the local matching method. The method can be applied to images corrupted by backscattering, taken under a non-homogeneous artificial light source. The authors, however, assumed that the depth in the supported window is constant, which led to wrong estimation of the scattering signal at depth discontinuities, especially in a highly non-homogeneous scattering signal area.

In this study, we propose a scattering removal technique, called *descattering*, followed by a standard stereo method where we focus on how to remove the scattering efficiently for stereo vision. The imaging model is derived in Section 2. From the model, the model-based descattering method is proposed, where we try to remove the scattering effect. The intermediate resulting images of the descattering method are defogged utilizing the well-known DCP [29]. Both steps above are shown in Section 3. The results of stereo vision of dense scattering scene of both synthetic images and real experimental images are shown in Section 4. The robot system and the accomplishment of a robot manipulation task are demonstrated in Section 5. Finally, Section 6 presents our conclusions.

2. Imaging Model

Three underlying assumptions are used in this approach:

- The illumination source is known and close to the cameras. This is feasible since the cameras and the light source are installed in the head of the robot.
- The scattering is single scattering. Although multiple scattering occurs, it is proven that utilizing single scattering model is effective in scattering removal [24,40,41].
- The input image I is given in the actual scene radiance values. The radiance maps can be recovered by inverting the acquisition response curve proposed by Debevec and Malik [42].

2.1. Single View Modeling in Scatterers Environment

Consider a vision system configuration in Figure 1. Let $\mathbf{X} = (X, Y, Z)$ and $\mathbf{x} = (x, y)$ be global coordinates of a point in space and its projection into image plane, respectively. $R_s(\mathbf{X})$ and $R_c(\mathbf{X})$ are the distances from a point in space \mathbf{X} to the light source and the left camera, respectively. $R_{c0}(\mathbf{x}_{obj})$ is the distance where the light field first intersects the line of sight (LOS), which is unique for every pixel \mathbf{x}_{obj} in the image. $I_s(\mathbf{X})$ is the irradiance of a point in space that is illuminated by the point light source and θ is the backscattering angle. B is the baseline. The measured intensity can be modeled as a linear combination of attenuated radiance $R(\mathbf{x}_{obj})$ (red line) (attenuated fraction of object radiance $L_{obj}(\mathbf{x}_{obj})$) and backscattering component $S(\mathbf{x}_{obj})$ (blue line) as follows:

$$I(\mathbf{x}_{obj}) = R(\mathbf{x}_{obj}) + S(\mathbf{x}_{obj}). \quad (1)$$

Note that the single scattering is assumed and the image blur due to the forward scattering [41] is not taken into account. The attenuated signal is:

$$R(\mathbf{x}_{\text{obj}}) = L_{\text{obj}}(\mathbf{x}_{\text{obj}})\tau(\mathbf{x}_{\text{obj}}) \quad (2)$$

where direct transmission is:

$$\tau(\mathbf{x}_{\text{obj}}) = e^{-cR_c(\mathbf{x}_{\text{obj}})} \quad (3)$$

where c is the attenuation coefficient (or extinction coefficient) of the environment due to absorption and scattering. The object radiance is given by:

$$L_{\text{obj}}(\mathbf{x}_{\text{obj}}) = I_s(\mathbf{X}_{\text{obj}})\rho(\mathbf{X}_{\text{obj}}) \quad (4)$$

where $\rho(\mathbf{x}_{\text{obj}})$ is the object reflectance. The irradiance of a point in space that illuminated by the point light source of intensity L_s is:

$$I_s(\mathbf{X}) = \frac{L_s Q(\mathbf{X})}{R_s^2(\mathbf{X})} e^{-cR_s(\mathbf{X})} \quad (5)$$

where $Q(\mathbf{X})$ expresses the non-uniformity of the illumination source. The falloff $\frac{1}{R_s^2(\mathbf{X})}$ is caused by free space light propagation. Since the baseline of the illuminator-camera is very small compared to the object distance, $R_s(\mathbf{X}) \approx R_c(\mathbf{X}) = \|\mathbf{X}\|$. Substituting Equations (3)–(5) into Equation (2), we obtain:

$$R(\mathbf{x}_{\text{obj}}) = \frac{L_s Q(\mathbf{X}_{\text{obj}})}{\|\mathbf{X}_{\text{obj}}\|^2} \rho(\mathbf{X}_{\text{obj}}) e^{-2c\|\mathbf{X}_{\text{obj}}\|}. \quad (6)$$

The total backscattering signal that the camera receives is:

$$S(\mathbf{x}_{\text{obj}}) = \int_{R_{c0}(\mathbf{x}_{\text{obj}})}^{R_c(\mathbf{x}_{\text{obj}})} b[\theta(\mathbf{X})] I_s(\mathbf{X}) e^{-cR_c(\mathbf{X})} dR_c(\mathbf{X}), \mathbf{X} \in \mathbf{X}_{\text{LOS}} \quad (7)$$

where $b[\theta(\mathbf{X})]$ is the phase function of backscattering. The LOS from the camera to object is:

$$\mathbf{X}_{\text{LOS}} = \left\{ \mathbf{X} : X = x_{\text{obj}} \frac{Z}{f}, Y = y_{\text{obj}} \frac{Z}{f}, Z \in [0, Z_{\text{obj}}] \right\} \quad (8)$$

where f is the cameras' focal length. To simplify the analysis, let us assume that $b[\theta(\mathbf{X}_{\text{LOS}})] \approx \tilde{b}$ is constant over the field of view, which is supported by [23,24], and that $Q(\mathbf{X}) \approx Q(\mathbf{X}_{\text{obj}})$ is constant along the LOS, which is supported by small camera-illuminator baselines. If there are several sources, Equation (7) applies to each source. Accumulating the integral for all sources yields the total backscattering. Equation (7) becomes:

$$S(\mathbf{x}_{\text{obj}}) = \tilde{b} L_s Q(\mathbf{X}_{\text{obj}}) \int_{R_{c0}(\mathbf{x}_{\text{obj}})}^{R_c(\mathbf{x}_{\text{obj}})} \frac{e^{-2cR_c(\mathbf{X})}}{R_c^2(\mathbf{X})} dR_c(\mathbf{X}), \mathbf{X} \in \mathbf{X}_{\text{LOS}}. \quad (9)$$

Tribitz and Schechner [23] derived the analytic solution of the integral in Equation (9) as Equation (10) and its approximation as Equation (11):

$$\frac{S(\mathbf{x}_{\text{obj}})}{\tilde{b} L_s Q(\mathbf{X}_{\text{obj}})} \approx \left\{ -\frac{e^{-2cR_c(\mathbf{X})}}{R_c(\mathbf{X})} - 2c \ln R_c(\mathbf{X}) - 2c \sum_{n=1}^{\infty} \frac{[-2cR_c(\mathbf{X})]^n}{n \cdot n!} \right\} \Bigg|_{R_c(\mathbf{X})=R_{c0}(\mathbf{x}_{\text{obj}})}^{R_c(\mathbf{X})=R_c(\mathbf{x}_{\text{obj}})}, \mathbf{X} \in \mathbf{X}_{\text{LOS}}, \quad (10)$$

$$S(\mathbf{x}_{\text{obj}}) \approx S_{\infty}(\mathbf{x}_{\text{obj}}) \left\{ 1 - e^{-k[R_c(\mathbf{x}_{\text{obj}}) - R_{c0}(\mathbf{x}_{\text{obj}})]} \right\} \quad (11)$$

where $S_{\infty}(\mathbf{x}_{\text{obj}}) \propto \tilde{b} L_s Q(\mathbf{X}_{\text{obj}})$ (considering Equation (10)) denotes the saturated backscattering value. It is worth noting that the non-uniformity of $S_{\infty}(\mathbf{x}_{\text{obj}})$ is attributed to the anisotropic pattern $Q(\mathbf{X}_{\text{obj}})$ in this special case. The constant parameter k depends on $R_{c0}(\mathbf{x}_{\text{obj}})$, c and \tilde{b} . From Equation (11), the rate

at which $S(\mathbf{x}_{obj}^i)$ increases with $\|\mathbf{X}_{obj}^i\|$ is set by parameter k . Since the baseline of illuminator-camera is very small compared to the object distance, in widefield lighting, we have $R_{c0}(\mathbf{x}_{obj}) \ll R_c(\mathbf{X}_{obj})$, thus $R_{c0}(\mathbf{x}_{obj}) \approx 0$. Substituting Equations (6) and (11) into Equation (1), noting that $R_c(\mathbf{X}_{obj}) = \|\mathbf{X}_{obj}\|$, the image's intensity becomes:

$$I(\mathbf{x}_{obj}) = \frac{L_s Q(\mathbf{X}_{obj})}{\|\mathbf{X}_{obj}\|^2} \rho(\mathbf{X}_{obj}) e^{-2c\|\mathbf{X}_{obj}\|} + S_\infty(\mathbf{x}_{obj})(1 - e^{-k\|\mathbf{X}_{obj}\|}). \quad (12)$$

Equation (12) resembles Koschmieder's law, which models daytime outdoor fog. The major difference is that in our case $S_\infty(\mathbf{x}_{obj})$ is spatial variant.

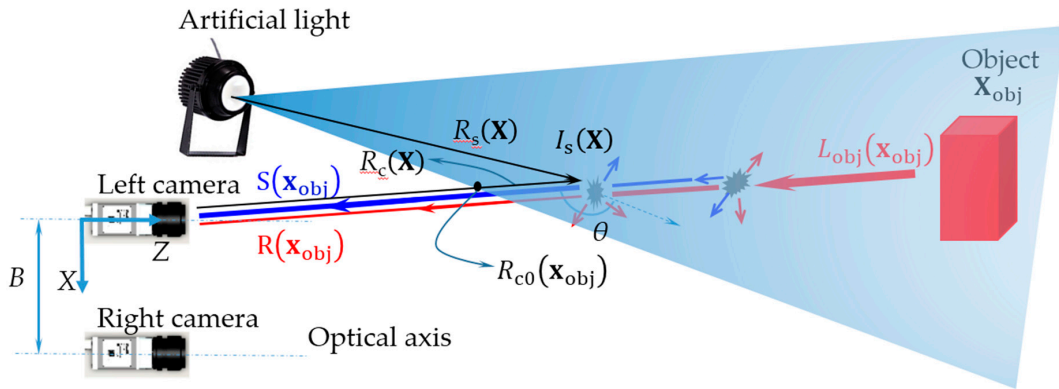


Figure 1. Stereo vision system configuration.

2.2. Stereo Modeling in Suspended Scatterer Environment

In a stereo vision system, using Equations (4) and (5), Equation (12) becomes:

$$I^i(\mathbf{x}_{obj}^i) = L_{obj}^i(\mathbf{x}_{obj}^i) e^{-c\|\mathbf{X}_{obj}^i\|} + S_\infty^i(\mathbf{x}_{obj}^i)(1 - e^{-k\|\mathbf{X}_{obj}^i\|}) \quad (13)$$

where \mathbf{X}_{obj}^i , $i = L, R$ are the coordinates of a point \mathbf{X}_{obj} in space with respect to left and right cameras, respectively. Noting that the global coordinates and left camera coordinates are the same ($\mathbf{X}_{obj}^L = \mathbf{X}_{obj}$), we have the relationship:

$$\mathbf{X}_{obj}^R = \mathbf{X}_{obj}^L - (B, 0, 0). \quad (14)$$

In Equation (13), the image coordinates of a point in space projected into the rectified left and right images are \mathbf{x}_{obj}^i , $i = L, R$. We also have:

$$\mathbf{x}_{obj}^R = \mathbf{x}_{obj}^L - [d(\mathbf{x}_{obj}^L), 0] \quad (15)$$

where $d(\mathbf{x}_{obj}^L)$ is the disparity map that pairs up corresponding pixels $\{\mathbf{x}_{obj}^L, \mathbf{x}_{obj}^R\}$.

In general, the system setup is more complicated than what was derived in Section 2.1: lighting geometry cannot be ignored, and there are several sources. In such cases, Tribitz and Schechner [23,24] show that the backscatter still follows the approximated model in Equation (11). However, $S_\infty^i(\mathbf{x}_{obj}^i)$ depends not only on the anisotropic pattern $Q(\mathbf{X}_{obj}^i)$ of the light source and the scattering parameters c and \tilde{b} , but also on the lighting geometry. The smaller camera-illuminator baseline is, the stronger the non-uniformity is. Equation (5) shows that the LOS that is closer to the light source receives stronger backscattering signal. The reason is that the irradiance $I_s(\mathbf{X})$ is very strong where the light field first meet the LOS. Thus, the two cameras sense a different backscattering signal, depending on their

geometric relationships to the light. That makes the stereo vision in scattering media more problematic because the intensity of the same object can be significantly different.

Figure 2a–c show the stereo pair of a clear scene, a foggy scene with natural light, and a foggy scene with an artificial light source, respectively. In the first example, since the images are taken in a clean environment, texture and contrast are preserved. Therefore, these images can be directly processed by utilizing conventional well-developed stereo vision algorithms. Figure 2b depicts the synthetic stereo images of the scene shown in Figure 2a in the presence of fog under natural light. In this case, the imaging model obeys Koschmieder’s law [25]. Due to attenuation, the greater the distance the signal propagates over, the weaker the object radiance that the cameras receive is. Thus, the contrast of these objects (inside the yellow rectangle) is low. Additionally, since the natural light is assumed to be parallel and uniform, the cameras capture the scattering signal, which depends on the air attenuation coefficient and the object distance. Although there are some difficulties in obtaining a depth map from these images due to poor contrast, the photo-consistency is held. Figure 2c represents an even more complicated case: the synthetic images of a scene under a foggy condition illuminated by an artificial light source that is installed under the two cameras. Besides suffering from poor contrast due to attenuation, the light adds a different scattering signal to the cameras, depending on lighting geometry. Consequently, the brightness of one object in two images (inside red rectangle) is not identical. Thus, the photo-consistency does not hold.

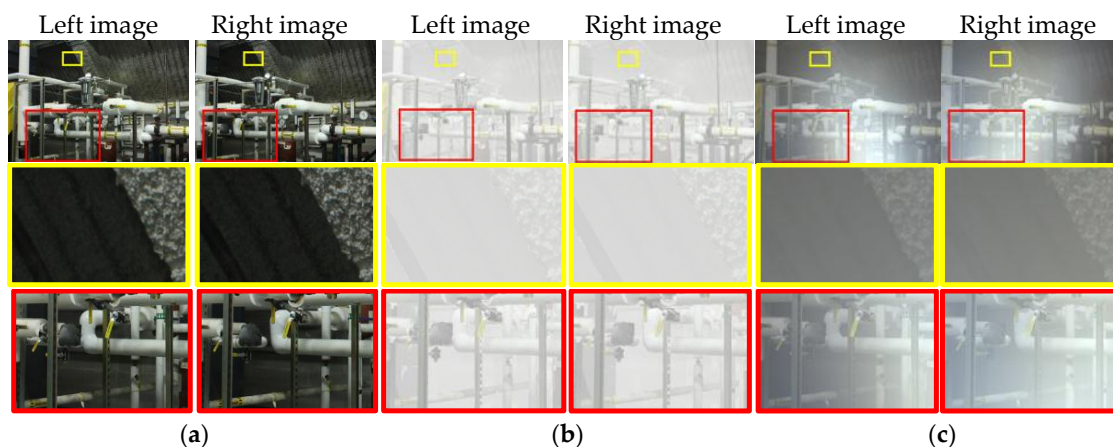


Figure 2. Example of stereo images (pipes from the Middlebury 2014 stereo datasets [9]) in different environment: (a) Stereo pair taken in clean environment; (b) Stereo pair taken in foggy environment with natural light source that suffers from attenuation uniform scattering; (c) Stereo pair taken in dense scatterer environment under an active light source that suffers both attenuation and non-uniform backscattering.

3. Backscattering and Fog Removal

3.1. Non-Uniform Backscattering Removal

3.1.1. Light Compensation

The first step is light compensation, which removes the non-uniformity of the backscattering. To do this, the measured image, modeled in Equation (13), is divided by the saturated backscattering signal, we obtain a light compensated image:

$$\hat{f}^i(\mathbf{x}_{\text{obj}}^i) = \hat{L}_{\text{obj}}^{d,i}(\mathbf{x}_{\text{obj}}^i) e^{-k\|\mathbf{x}_{\text{obj}}^i\|} + (1 - e^{-k\|\mathbf{x}_{\text{obj}}^i\|}) \quad (16)$$

where the distorted radiance of object is defined as:

$$\hat{L}_{\text{obj}}^{\text{d},i}(\mathbf{x}_{\text{obj}}^i) = \frac{L_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i)}{S_{\infty}^i(\mathbf{x}_{\text{obj}}^i)} e^{-(c-k)\|\mathbf{x}_{\text{obj}}^i\|} \quad (17)$$

where $L_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i)/S_{\infty}^i(\mathbf{x}_{\text{obj}}^i)$ is a spatial-varying value. It depends on the geometric configuration of the light. This means that applying the light compensation step results in many local radiometric differences in the object signal. However, it will be compensated after defogging. The light compensated image in Equation (16) is similar to Koschmieder's law with the airlight equal 1. Let us denote $\tau^{k,i}(\mathbf{x}_{\text{obj}}^i) = e^{-k\|\mathbf{x}_{\text{obj}}^i\|}$, which is the modified direct transmission.

Noting that $\hat{L}_{\text{obj}}^{\text{d},i}(\mathbf{x}_{\text{obj}}^i)$ in Equation (17) is neither the reflectivity nor radiance of the object. It, however, is an enhanced image, with radiometric distortion, from the original corrupted image by strong backscattering.

3.1.2. Saturated Backscattering Estimation

Based on (11), saturated backscattering can be easily estimated:

$$\lim_{\|\mathbf{x}_{\text{obj}}^i\| \rightarrow \infty} \hat{I}^i(\mathbf{x}_{\text{obj}}^i) = S_{\infty}^i(\mathbf{x}_{\text{obj}}^i). \quad (18)$$

Thus, saturated backscattering can be pre-calibrated by taking the void images where there is no object in the images ($\|\mathbf{x}_{\text{obj}}^i\| \rightarrow \infty$).

However, in our experiment, due to space limitation, we took pictures of very dense steam and fog (e.g., $c \approx 2.3 \text{ m}^{-1}$) scenes where no object can be seen. As derived in Section 2.2, saturated backscattering depends on the attenuation coefficient c . However, from Equation (10), we can obtain the following relationship:

$$S_{\infty}^i(\mathbf{x}_{\text{obj}}^i, c < 2.3) \approx k_c S_{\infty}^i(\mathbf{x}_{\text{obj}}^i, c = 2.3) \quad (19)$$

where $k_c < 1$ is the constant gain, which depends on the attenuation coefficient. The constancy of k_c at the specific attenuation coefficient was confirmed by our experiment, for example, $k_{c=1.5} = 0.86 \pm 0.05$.

Figure 3 illustrates the saturated backscattering signal of two different system configurations. The images are the original images without any color correction. In the first setup, the light was put under the cameras, and steam was generated by the steam generator using pure water. The light was placed above the cameras in the second setup, and the fog was produced by a fog machine using oil.

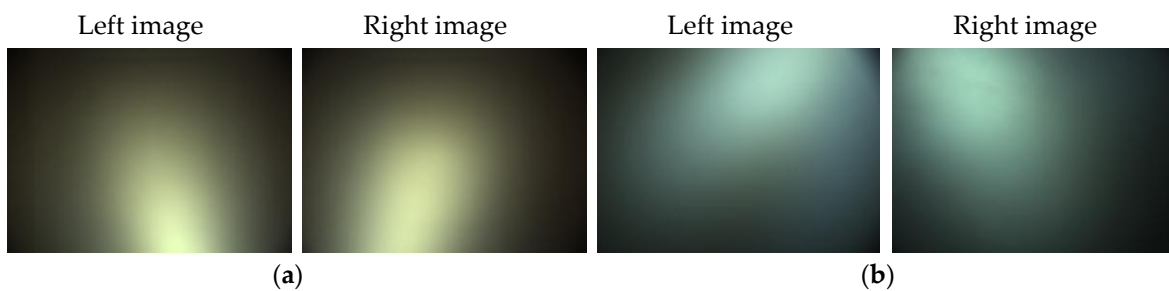


Figure 3. An example of saturated backscattering: (a) Lighting setup 1; (b) Lighting setup 2.

3.2. Defogging

3.2.1. DCP-Based Defogging

DCP [29] is employed to remove the fog, a process called defogging, of the light compensated image in Equation (16). Let us summarize the DCP proposed by He et al. [29]. The dark channel of the light compensated image is defined as:

$$\tilde{J}^d(\mathbf{x}_0) = \min_{\mathbf{x} \in \Omega(\mathbf{x}_0)} \left[\min_{c \in \{r, g, b\}} \hat{J}^c(\mathbf{x}) \right] \quad (20)$$

where $\Omega(\mathbf{x}_0)$ is the local patch centered at \mathbf{x}_0 . The patch transmission is then calculated as:

$$\tilde{\tau}^k(\mathbf{x}) = 1 - \tilde{J}^d(\mathbf{x}). \quad (21)$$

Different from the original DCP method, we employ guided image filtering [17] to refine the raw transmission map in Equation (21) in order to obtain $\tau^k(\mathbf{x})$. The distorted object radiance can be obtained by inverting Equation (16):

$$\hat{L}_{\text{obj}}^{\text{d},i}(\mathbf{x}_{\text{obj}}^i) = \frac{\hat{J}(\mathbf{x}_{\text{obj}}^i) - 1}{\max[\tau^{k,i}(\mathbf{x}_{\text{obj}}^i), \tau_0]} + 1. \quad (22)$$

Transmission can be very close to zero; thus, it is restricted to the lower bound τ_0 . There is radiometric distortion in the distorted object radiance, as shown in Equation (17). Therefore, to preserve the photo-consistency in the left and right images, the radiometric distortion must be eliminated. This can be done easily by multiplying the distorted object radiance by saturated backscattering $S_{\infty}^i(\mathbf{x}_{\text{obj}}^i)$ to obtain the modified object radiance as follows:

$$\hat{L}_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i) = L_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i) e^{-(c-k)\|\mathbf{x}_{\text{obj}}^i\|}. \quad (23)$$

It is also worth noting that $\hat{L}_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i)$ is not the original radiance of the object. The value $e^{-(c-k)\|\mathbf{x}_{\text{obj}}^i\|}$ either attenuates (when $-(c-k) < 0$) or amplifies (when $-(c-k) > 0$) the original object radiance. However, in our experiment, the modified radiance images are useful for both reviewing the scene and reconstructing depth map. For simplicity, we call $\hat{L}_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i)$ a *defogged image* in our paper.

DCP was designed for natural images. The assumption may not hold for indoor human-made scenes. The main reason is that DCP can detect the specular reflection [43]. By utilizing the active polarization system [24] (explained in Appendix A), the specular reflection can be removed; thus, we verify that the DCP works properly in our system.

3.2.2. Normalization-Based Image Correction

From our observation, when the fog is very dense and uniform, the modified direct transmission $\tau^k(\mathbf{x}_{\text{obj}})$ is almost a constant, which is very small; thus, the backscatter $S^i(\mathbf{x}_{\text{obj}})$ is close to its saturation $S_{\infty}^i(\mathbf{x}_{\text{obj}})$. Consequently, the minimum intensity of the light compensated image is set by atmospheric veil $[1 - \tau^k(\mathbf{x}_{\text{obj}})]$. Therefore, by normalizing the light compensated image $\hat{J}^i(\mathbf{x}_{\text{obj}}^i)$, we can efficiently both remove the atmospheric veil and scale $\hat{L}_{\text{obj}}^{\text{d},i}(\mathbf{x}_{\text{obj}}^i)$ to $[0, 1]$. The normalization image is defined as follows:

$$\hat{J}_n^i(\mathbf{x}_{\text{obj}}^i) = \frac{\hat{J}^i(\mathbf{x}_{\text{obj}}^i) - \min[\hat{J}^i(\mathbf{x}_{\text{obj}}^i)]}{\max[\hat{J}^i(\mathbf{x}_{\text{obj}}^i)] - \min[\hat{J}^i(\mathbf{x}_{\text{obj}}^i)]}. \quad (24)$$

The image is an approximation of $\hat{L}_{\text{obj}}^{\text{d},i}(\mathbf{x}_{\text{obj}}^i)e^{-k\|\mathbf{x}_{\text{obj}}^i\|}$ in Equation (16). Then, to remove radiometric distortion, we define the compensated normalization image as:

$$\hat{R}_n^i(\mathbf{x}_{\text{obj}}^i) = \hat{J}_n^i(\mathbf{x}_{\text{obj}}^i)S_{\infty}^i(\mathbf{x}_{\text{obj}}^i). \quad (25)$$

Only scattering removal is involved in this method. The attenuation was not removed in this image; thus, it still suffers from poor contrast. From that physical meaning, we call $\hat{R}_n^i(\mathbf{x}_{\text{obj}}^i)$ a descattered image. However, we will show in Section 4 that this method is feasible for stereo vision in uniform steam environments. However, it fails in the case of non-uniform steam.

Figure 4 shows our descattered and defogged results. The first row shows images when the fog is uniform while the second row depicts the images in the case of non-uniform fog. The image in Figure 4a was taken in a very dense fog environment ($c = 1.6 \text{ m}^{-1}$) associated with lighting setup 2 in Figure 3. In Figure 4b, the light compensated image $\hat{J}^i(\mathbf{x}_{\text{obj}}^i)$ of the input image is illustrated, which were scaled into [0,1] for visualization. Figure 4c,d show descattered and defogged results from the proposed method, respectively. Figure 4e,f are nighttime dehazed results of Zhang et al. [35] and Li et al. [36], respectively. In the case of uniform fog, it can be seen that both descattered and defogged images from our method are better than that of [35,36].

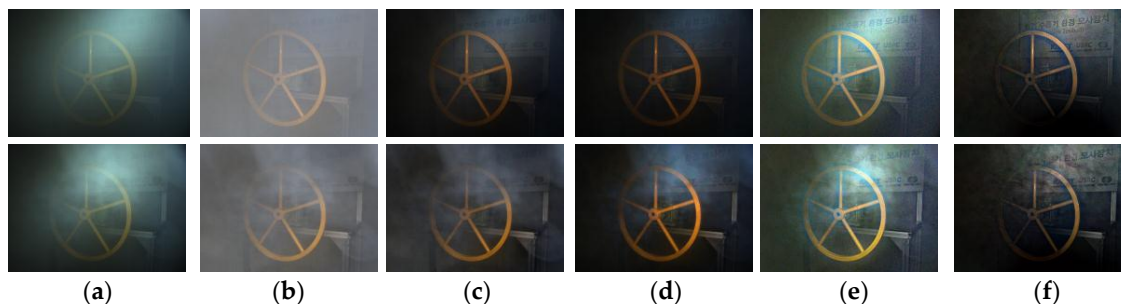


Figure 4. Descattering and defogging result. (a) Corrupted images; (b) Light compensated image; (c) Our descattered result; (d) Our defogged result; (e) Nighttime dehazing from Zhang et al. [35]; (f) Nighttime dehazing from Li et al. [36].

The method of [35] is incapable of removing image glow whereas in the result of [36], the dark area becomes very dark. In the case of non-uniform fog, our defogged method and the method in [36] show better ability of non-uniform fog removal. The result of [36], however, still makes the dark area become darker.

4. Stereo Vision Results

4.1. Experimental Setup

- In the first setting, the stereo baseline is 10 cm. The light is put under the cameras. The light source and cameras are not coaxial. The experiment was conducted in a booth with dimensions of $3 \times 1.5 \times 1.6 \text{ m}^3$. We utilized a steam generator to generate the steam using pure water inside the cabin. The generated steam's temperature is 100–120 °C. Our system is able to produce steam as dense as an attenuation coefficient of 1.15 m^{-1} .
- In the second setup, the stereo vision is the same as the previous configuration. However, the light source is placed above the cameras and coaxial to cameras. This experiment was done in a room with dimensions of $6 \times 4 \times 2.5 \text{ m}^3$. To generate fog in such a big room, we utilized a fog machine (CHAMP-1500W, Joongang Special Lights, Seoul, Korea) that uses oil.

We make use of visibility to estimate the steam and fog density. The visibility is a measure of distance at which an object can be clearly discerned from the background. Visibility V is calculated as:

$$V = \frac{C}{c} \quad (26)$$

where C is a constant depending on contrast ratios. Contrast ratios are between 0.018 and 0.03. A contrast ratio of 0.02 is usually used to calculate the visual range; thus, $C = 3.912$. The attenuation coefficient is calculated as follows:

$$c = \frac{1}{L} \ln \frac{I}{I_0} \quad (27)$$

where L is the distance that the light travels from the source to the receiver. I_0 and I are the intensity measured when light travels in the clear condition and the foggy condition, respectively. To measure the attenuation coefficient c and then visibility V , a HeNe laser (wavelength of 632.8 nm and power of 0.8 mW) and a photodiode sensor (S120C), both from Thorlabs, Newton, NJ, USA, were employed as an emitter and receiver, respectively. It should be noted that the attenuation coefficient c is wavelength-dependent. The longer the wavelength is, the higher the attenuated coefficient c is.

4.2. Stereo Results from Synthetic Images

Twelve datasets (Middlebury 2014 stereo datasets) from [9] were selected and used to generate synthetic data. The images were resized by half. We created synthetic images based on our imaging model derived in Section 2 with the provided ground truth disparity map. We normalized and scaled the ground truth depth map into a range from 0.5 m to 2.5 m. In the attenuated signal term, the non-uniformity of the illumination source is negligible. Only the attenuation of object radiance (from the original images) is considered. A backscattering signal is added to images based on our real pre-calibrated saturated backscattering signal $S_{\infty}^i(\mathbf{x}_{\text{obj}}^i)$.

The criteria to evaluate the quality of the disparity map from the synthetic image is the percentage of good matching pixels [8]. The threshold value of one was used. If the difference between the estimated disparity and the ground truth is larger than one, the pixel is considered to be a bad pixel. Otherwise, it is a good pixel.

We found that our descattered images $\hat{R}_n^i(\mathbf{x}_{\text{obj}}^i)$, derived in Section 3.2.2, without DCP-based defogging provide a better stereo result in the case of dense uniform steam. However, for images of non-uniform steam scenes, the defogged images $\hat{L}_{\text{obj}}^i(\mathbf{x}_{\text{obj}}^i)$, derived in Section 3.2.1, work better. The reason is that the defogging algorithm based on DCP is based on statistics; thus, the estimation of transmission may not be accurate. Therefore, the color, which is very sensitive to the transmission map, in left and right images is less similar after defogging, which causes wrong matching. The descattered image, on the other hand, is very close to the modified object radiance. The reason is that the modified transmission $\tau^k(\mathbf{x}_{\text{obj}})$ is almost constant and close to 1 in a dense scatterer environment. In the case of non-uniform fog or steam, because c and k are spatial-varying, the above assumption does not hold. In this case, DCP based defogging can remove the non-uniformity of the fog in the image; thus, the stereo vision quality of defogged images is better than that of descattered images. This will be proven in both synthetic images in this section and real images in the next section.

Semi-global matching (SGM) [44] was employed as a stereo vision algorithm in our real robot manipulation task. Table 1 shows a comparison of the disparity map quality between the descattered and defogged images of two kinds of conditions, namely, uniform steam ($V = 3$ m) and non-uniform steam ($V \in [3, 4]$ m). When dealing with images corrupted by uniform dense steam, descattered images are about 10% better than defogged images. In the case of non-uniform steam, defogged images, however, provide a 7% better result. Thus, the choice of making use of descattered images or defogged images depends on whether the environment is uniform.

For evaluation, we compared the disparity map from our descattering and defogging method with those of backscatter-corrupted images, Negahdaripour and Sarafranz [40], Zhang et al. [35], and Li et al. [36]. The method in [40] improves stereo matching by incorporating backscattering cues. This method is a local matching method and can obtain the depth map directly. The authors utilized

Normalized Sum of Square Difference (NSSD) with the mean subtraction function as the matching cost. The nighttime dehazing methods in [35,36] can improve the visibility of a hazed image of a scene illuminated by active light sources. We implemented the method in [40] and ours using Matlab, while the authors of [35,36] provided their software run in C and Matlab, respectively. We can freely choose the stereo algorithm to process our descattered and defogged images. However, since the method in [40] is based on NSSD, we treat the other images in the stereo vision step using the same matching cost function for a fair comparison. It should be noted that in our robot manipulation, we employed SGM.

Table 1. Correct matching rate from proposed descattered and defogged images.

Dataset Name	Uniform Fog		Non-Uniform Fog	
	Descat (%)	Defog (%)	Descat (%)	Defog (%)
Adirondack	67.10	52.79	30.84	42.73
Backpack	76.69	72.13	43.84	51.1
Cable	61.72	40.37	15.90	19.33
Classroom1	85.87	64.49	10.04	24.79
Flowers	44.82	48.63	19.04	14.56
Motorcycle	76.22	71.86	43.32	58.35
Pipes	66.96	58.78	49.19	49.68
Recycle	67.23	53.2	15.59	25.43
Shelves	47.45	40.44	24.88	35.63
Storage	61.77	54.75	33.81	33.21
Sword1	77.84	68.87	39.73	50.75
Sword2	42.21	27.99	6.85	11.82
Average	64.66	54.53	27.75	34.78

Table 2 illustrates the summarized comparison of the stereo vision results using NSSD in three conditions, namely, lighting setups 1 and 2 with uniform fog, and lighting setup 1 with non-uniform fog. The data are the average correct rate of the 12 datasets. In the case of uniform fog, our descattered images were used for stereo vision. In lighting 1, the proposed method shows at least a 14% higher correct rate than all the other methods. The stereo results obtained from corrupted images, dehazed images using the method in [36], and the stereo results obtained by using method in [40] are almost identical while the stereo results obtained from dehazed images using the method in [35] are worse than using corrupted images. There are several reasons for this. First, NSSD is capable of compensating offset and gain [45]; thus, it already works well in the case of corrupted images. As mentioned in Section 1, the method in [40] assumed that the depth in the supported window is constant, which led to wrong estimation of the scattering signal at depth discontinuities, especially in a highly non-homogeneous scattering signal area. In the datasets with lighting 1, there is strong backscattering at the high depth discontinuities areas of the datasets, as in the example of the Pipes dataset shown in Figure 5. Therefore, there is no improvement compared with the corrupted images. The method in [35] provide the worst results because this method is unable to remove the strong backscatter in the image due to their imaging model. The method in [36] has the ability to remove glow, and hence works better than that in [35]. In lighting 2, as shown in Figure 6, the light illuminates the scene above the camera; thus, a strong backscattering signal projects into the higher area of images. In these datasets, these regions have fewer depth discontinuities. Consequently, the disparity map correct rate obtained by using the method in [40] is about 11% greater than that of the original corrupted images. The nighttime dehazing methods in [35,36], and our method show the identical correct rate compared with the rate in the previous case. It should be noted that our disparity map quality is the best and is 20% higher than the disparity obtained from the input images. In the case of non-uniform steam, the results of dehazed images from [36] and our defogged images have almost the same quality and slightly higher quality than the others.

Table 2. Evaluation of stereo vision from scatter-corrupted images; the stereo vision method is NSSD.

Lighting	Corrupted Image (%)	[40] (%)	[35] (%)	[36] (%)	Proposed Method (%)
Setup 1—uniform	33.12	33.03	25.12	34.30	47.84
Setup 2—uniform	26.39	37.29	23.04	32.82	46.16
Setup 1—non-uniform	23.70	19.37	22.25	25.70	25.99

Since in the real system we employ SGM, the proposed method is also compared with backscatter-corrupted images [35,36], using SGM as the stereo algorithm, as shown in Table 3 and an example in Figure 6. In this case, SGM performs worse than NSSD when using corrupted images while it performs better using dehazed images from [35,36], and ours. When using SGM, the method in [35] provides slightly better quality than the original images. In the case of uniform fog, the proposed method improves the matching rate by about 35% and 20% compared with input images and dehazed image obtained by using the method in [36], respectively. In the case of non-uniform steam, our method and the method in [36] are nearly the same, being 10% greater than the inputs.

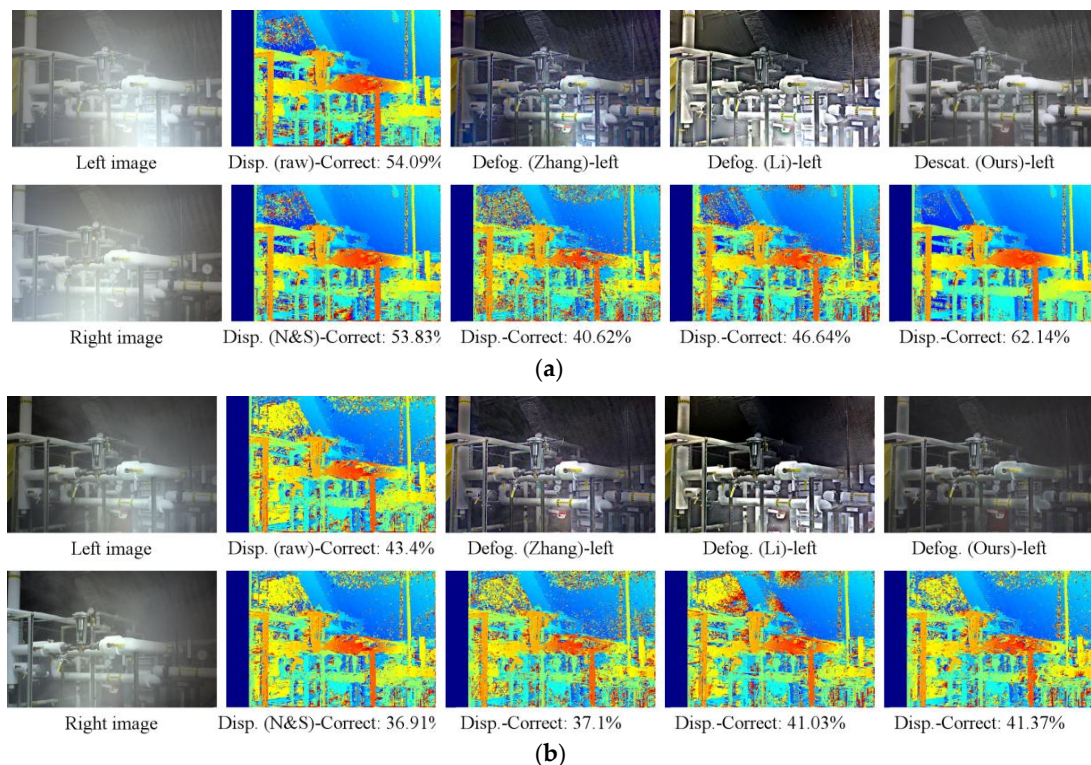


Figure 5. An example of synthetic images of Pipes [9]; the stereo method is NSSD: (a) Lighting 1—uniform; (b) Lighting 1—non-uniform. The first column is corrupted images. The second column shows the disparity map from input images and the one obtained by using the method in [40]. “N&S” stands for Negahdaripour and Sarafraz [40]. The third to last columns are the defogged (or descattered) images and disparity maps using the methods from [35,36] and the proposed method, respectively. “Disp.” and “Defog.” stand for disparity map and defogged image, respectively.

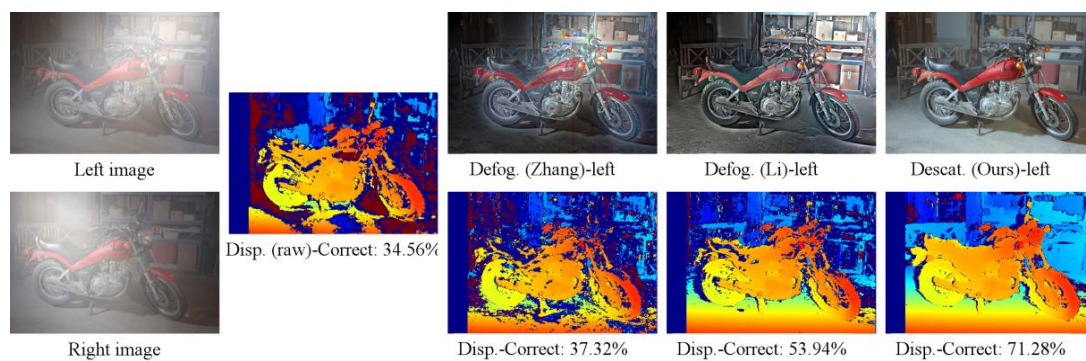


Figure 6. An example of synthetic images of Motor in the case of lighting 2 and uniform fog: The first column is corrupted images; the second is disparity from input images; the third to the last columns are the defogged (or descattered) images and disparity maps obtained by using the methods in [35,36] and the proposed method, respectively.

Table 3. Evaluation of stereo vision from scatter-corrupted images; the stereo vision method is SGM.

Lighting	Corrupted Image (%)	[35] (%)	[36] (%)	Proposed Method (%)
Setup 1—uniform	28.09	31.31	45.46	64.66
Setup 2—uniform	19.49	26.99	41.89	55.53
Setup 1—non-uniform	24.94	29.14	34.57	34.78

4.3. Stereo Vision Results from Real Images

In Section 2, it is assumed that the input image $I^i(x_{obj}^i)$ is given in the actual scene radiance values. The radiance maps can be recovered by inverting the acquisition response curve proposed by Debevec and Malik [42]. This is the only preprocessing step, which is employed in our experiment. This step also helps reducing variations in color which are produced by two different cameras in the stereo vision system.

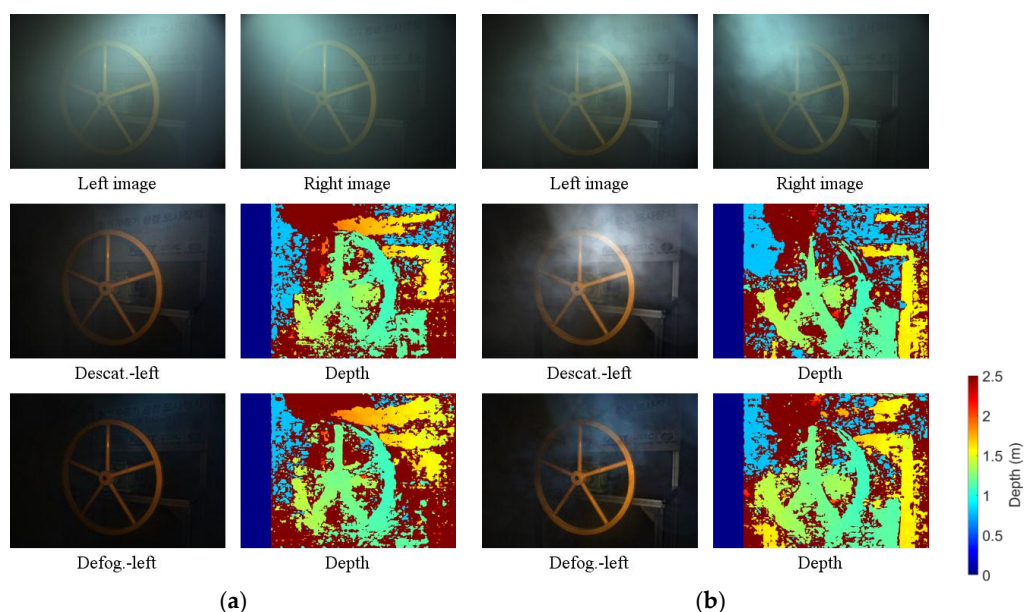


Figure 7. Experimental results; the stereo method is SGM: (a) Lighting 2—uniform fog of $V = 2.4$ m; (b) Lighting 2—non-uniform fog. The first row is corrupted images; the second and third row are descattered and defogged images, respectively.

Figure 7 shows a comparison of the depth map quality between the descattered and defogged images from the proposed method of two kinds of conditions, namely, uniform ($V = 2.4$ m) and non-uniform steam. When dealing with images corrupted by uniform dense steam, descattered images are better than defogged images. In the case of non-uniform steam, defogged images, however, provide better result. This is consistent with the simulation results as shown in Table 1.

We depicted several real experiment data in Figures 8 and 9. Figure 8a,b show two examples of lighting setup 1 in dense uniform steam (V are 4.24 and 3.39 m) using NSSD. In Figure 8a, the proposed method performs the best and more depth detail can be reconstructed while [40] shows the worst result in reconstructing the chair. The reason for this is the assumption of [40] as mentioned in the previous section. The method in [40], however, has better ability to estimate the background depth. Figure 8b shows a similar trend. Figure 8c,d illustrate examples of non-uniform fog under setup 2 using NSSD. In both cases, the valve is tilted at an angle of 20° to 30° compared with cameras' optical axis and the distance from the center of the valve to cameras is 1.2 m. In both cases, the proposed method outperforms the input images, [35,36,40], in constructing the depth of object, especially the valve. The depth results from input images and that obtained by using method in [40] are the worst in both cases, especially in strong backscattering regions. In Figure 8d, the method in [35] performs better than that in [36] because the dehazed images of [36] are very dark in the lower areas.

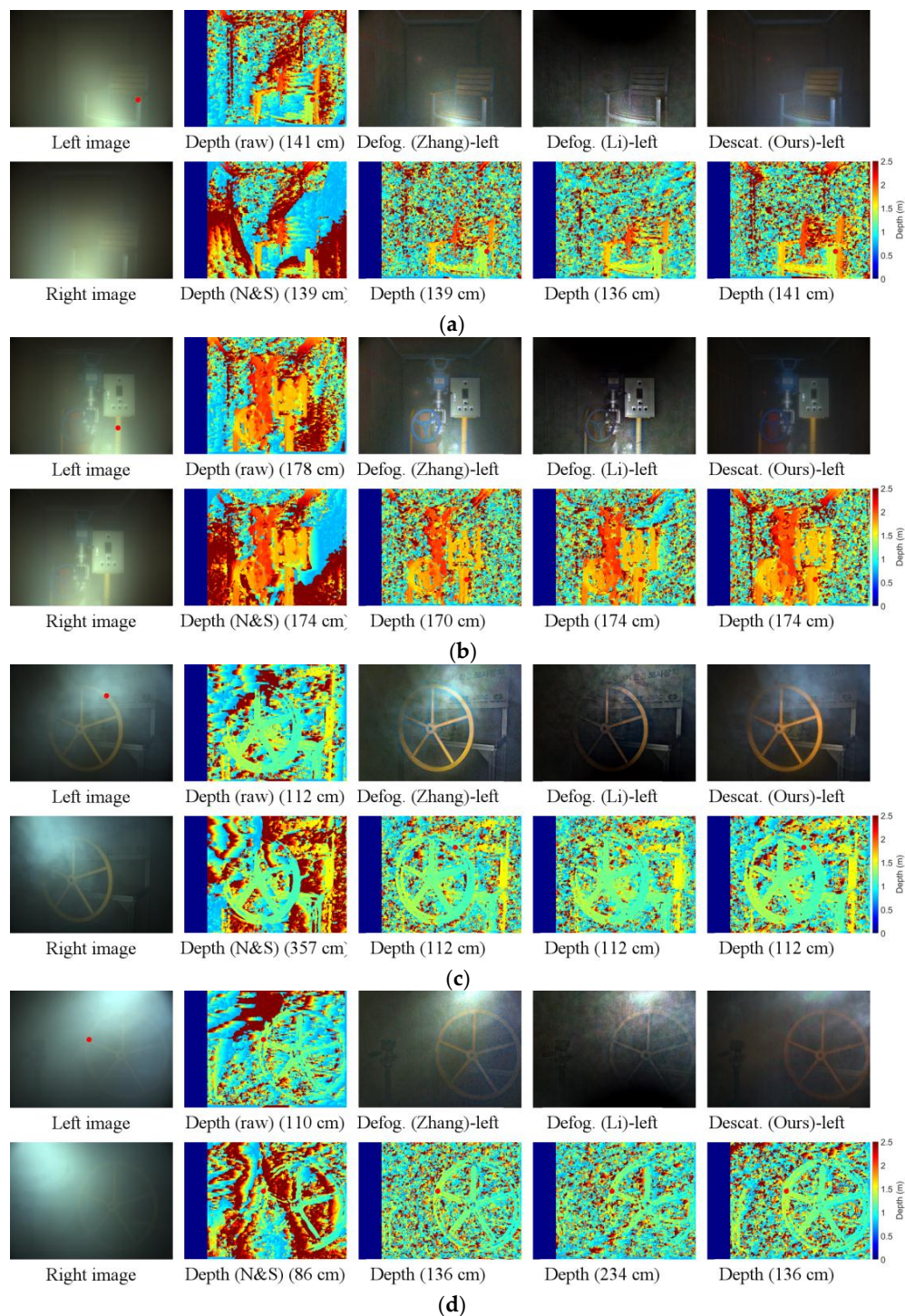


Figure 8. Experimental results; the stereo method is NSSD: (a) Lighting 1— $V = 4.24$ m; (b) Lighting 1— $V = 3.39$ m; (c,d) Lighting 2—non-uniform. The first column is corrupted images; the second column shows the depth maps from input images and those obtained by using method in [40]; the third to last columns are the defogged (or descattered) images and disparity maps using methods in [35,36] and the proposed method, respectively. The number under the every depth map is the measured depth at the red dot.

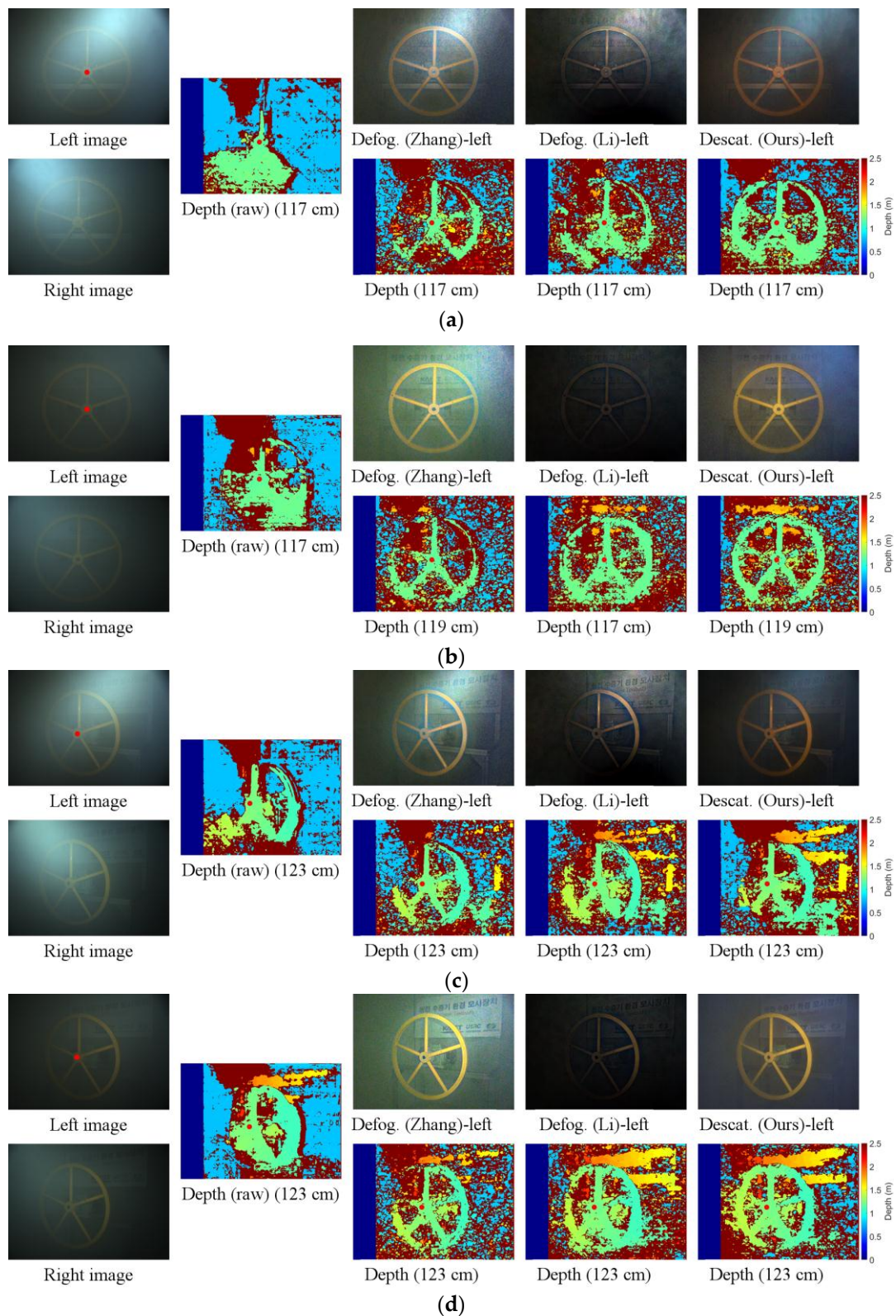


Figure 9. Experimental results; the stereo method is SGM: (a) Lighting 2— $V = 1.71$ m and polarization angle of 45° ; (b) Same as (a) with polarization angle of 90° ; (c) Lighting 2— $V = 2.39$ m and polarization angle of 45° ; (d) Same as (c) with polarization angle of 90° . The first column is corrupted images; the second is disparity from input images; the third to last columns are the defogged (or descattered) images and disparity maps obtained using methods in [35,36] and the proposed method, respectively. The number under the every depth map is the measured depth at the red dot.

Figure 9 depicts examples under setup 2 using SGM and the effect of polarization. In Appendix A, we discuss about the active polarization lighting and the effects of polarization. Figure 9a,c show two examples of lighting setup 2 in dense uniform steam (V are 1.71 and 2.39 m) when the polarization angle is 45° . In both cases, the distance from the center of the valve to cameras is 1.2 m. Figure 9b,d show data under the same conditions as Figure 9a,c, respectively, when the polarization angles are 90° . In both cases, the proposed method outperforms the input images [35,36], in reconstructing the depth of the object, especially the valve.

For every method, utilizing orthogonal polarization provides a better result than using a polarization angle of 45° . Directly using input images does not work well in both polarization angles. One important observation is that all methods can estimate the distance to the center of the valve accurately. Our system is better since it provides more constructed points. Finally, another crucial factor to utilize the vision algorithm in a real robot application is real-time capability. Table 4 shows the processing time to obtain the descattered or defogged images. We took the average processing time when processing 100 images continuously. The software and code run in different environments. Authors of [35] provided their software, which is an executable file in C++ environment, while authors of [36] provided a protected function run in Matlab. We implemented our descattering and defogging method using Matlab (non-optimized implementation). Thus, this is not a fair comparison. Nevertheless, we demonstrate a near real-time capability of our descattering method to enhance the input images for the stereo vision system with a processing time of 34 ms for a single image.

Table 4. Processing time.

Resolution	Zhang et al. [35] (ms)	Li et al. [36] (ms)	Ours (ms)	
			Descat	Defog
780×580	17,470	20,520	34	860

5. Verification with Robot Manipulation

To verify the proposed algorithm, we successfully demonstrated robot manipulation in a foggy condition. In this chapter, the robot system of the manipulator is introduced, and the results of a valve turning mission in a foggy condition are presented.

5.1. The Robot System of the Manipulator

The robot manipulator is constructed with seven actuators (shoulder: three axes, elbow: one axis, and wrist: three axes) to mimic the human arm configuration, which is a redundant system. The actuator models used in the robot manipulator are PRL+120, ERB-145, and ERB-115, which are produced by SCHUNK Corporation (Mengen, Germany). The specifications of the actuator model are given in Table 5.

Table 5. Specifications of the actuators.

Specification	Unit	ERB-115	ERB-145	PRL-120
Max Speed	$^\circ/s$	72	72	25
Nominal Torque	Nm	7	35	216
Max Torque	Nm	19	64	372
Max rotation angle	$^\circ$	340	340	360
Weight	kg	1.8	3.9	3.6

5.2. Manipulation Experiment in Foggy Condition

We performed a manipulation experiment in foggy conditions to verify the effectiveness of the descattering method in a real robotics application.

5.2.1. Experiment Environment

The experiment environment is illustrated in Figure 10. The LiDAR (MultiSense SL from Carnegie Robotics, Pittsburgh, PA, USA) is also placed in the experiment environment for comparison. With the laser-based visibility measurement system, we monitor the visibility. To generate the fog, we used the fog machine, which has a power of 1500 W.

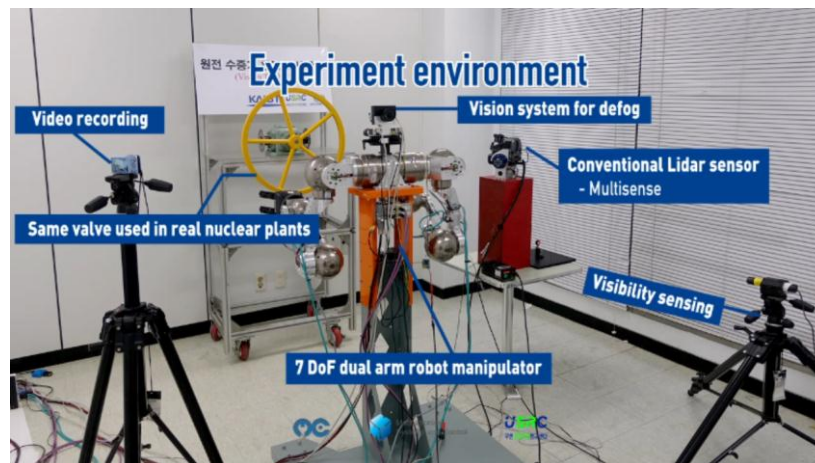


Figure 10. Experiment environment of valve turning manipulation in foggy condition.

With the fog machine, the foggy condition where the visibility range is under 2 m can be generated in experimental setup 2, as explained in the previous section. As seen in Figure 11, the LiDAR works well in a clear environment. However, in the dense fog condition, it is unable to work.

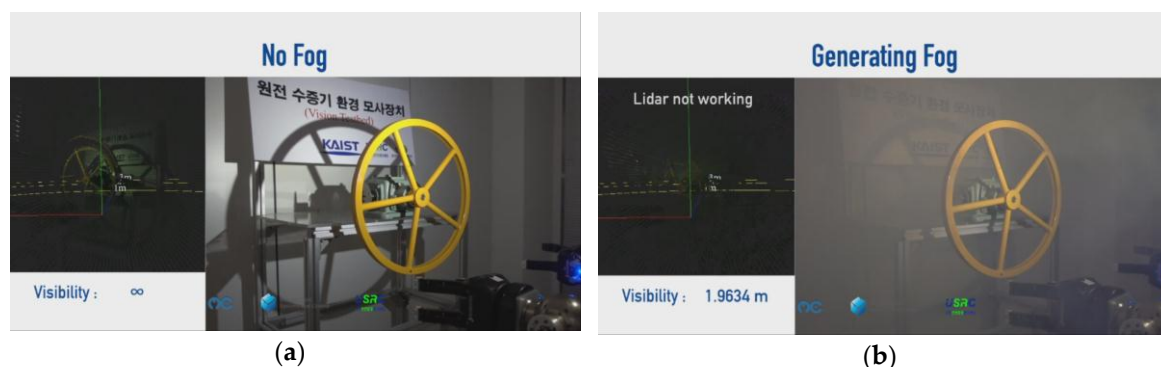


Figure 11. Visibility comparison: (a) without fog and (b) dense fog.

5.2.2. Experiment Results

With the proposed descattering-then-stereo algorithm, we are able to obtain a depth map. Based on the depth map, points of the valve are manually selected by the user. From these points (for example, 10 points), the center coordinate, normal vector, and radius of the valve is accurately extracted in a foggy condition. As shown in Figure 12, the obtained radius, the position of the center, and the normal vector of the valve are 31.54 cm, (70.82, 2.15, 4.28 cm), and (1.00, 0.03, -0.05), respectively.

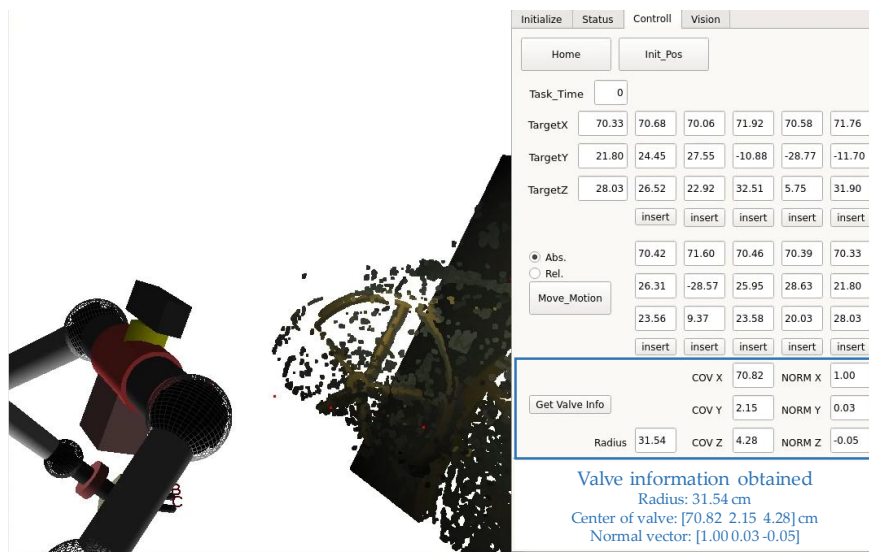


Figure 12. The obtained valve information.

With the valve information, the mission to turn the valve is successfully performed, as shown in Figure 13. The operator controls the robot remotely only using the vision data. As shown in Figure 9, backscatter-corrupted images generate poor quality depth maps. Therefore, although our method does not directly benefit the manipulation task, it helps providing higher quality input images for stereo vision. More specifically, our method reconstructs denser depth maps, from which we can select more points from a larger variety of positions to produce a more accurate estimation.



Figure 13. The snapshot of the robot turning the valve in dense fog condition.

6. Conclusions

In this paper, we present our descattering method, which can enhance images corrupted by strong non-uniform backscattering from an active illumination source. The method is very promising since it can enhance images for stereo vision and it is near real-time capable.

It is worth noting that our method is a model-based method. The proposed method and method from [40] are based on the pre-calibrated saturated backscattering. Thus, it is not surprising that our method outperforms the methods from [35,36]. However, we have proposed a simple method that is able to enhance the images of dense fog or dense steam scenes very efficiently for stereo vision.

The method is not restricted to our application. It can be utilized in other applications where active lighting is necessary, such as underwater robots.

An important issue in using our method is the choice whether to use descattered images or defogged images, such that a uniform fog/steam environment requires descattered images while a non-uniform environment requires defogged images. In practical operation, as mentioned in Section 5, the operator controls the robot remotely using vision data. The operator is also the one to make this decision. Algorithm to automatically detect non-uniform (heterogeneous) fog environment would be an issue for future works.

Acknowledgments: This research was supported by the KHNP (Korea Hydro & Nuclear Power Co., Ltd.). Funds for covering the costs to publish in open access was supported by the Department of Mechanical Engineering at KAIST under the BK21 Plus program.

Author Contributions: C.D.T.N. derived the imaging model and developed algorithm for descattering. J.P. designed and made the robot manipulation system. K.Y.C made the software. C.D.T.N., J.P., and K.Y.C. designed and performed experiments. K.-S.K. and S.K. discussed the weaknesses of the system while it was being implemented and test. C.D.T.N wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Polarization-Based Backscattering Removal

Using artificial light that is close to the camera, as in our system, causes a strong backscatter. To partly remove the strong backscattering, we utilize polarization, which has proven excellent ability to reduce backscattering [23,24,46–48]. We simply modify the conventional image acquisition system to make active polarization, similar to [23,24], by adding three polarizers: each one is mounted in front of the light source and every camera in the stereo vision system. To ensure photo-consistency, we aligned the polarizers in front of the two cameras so that they are in the same state. We use polarization angle to call the angle between the state of the polarizer of the light source and the state of the polarizers of the two cameras. This system provides a pure optics-based scattering removal for improving visibility. The best situation for backscattering removal is when the polarization angle is 90° [23,24]. Through our experiment, we found that it is very feasible to utilize the active polarizer. The active polarization lighting reduces both saturated backscattering and the constant parameter k .

To estimate the parameter k , we first apply the initial step of our proposed descattering method, as in Equation (16). The feature extraction and matching are then done by employing speeded up robust features (SURF) [49]. It should be noticed that SURF is unable to detect any feature when applied to the original corrupted images. The constant k can be estimated as:

$$k = \frac{\sum_{j=1}^N k_j}{N} \quad (\text{A1})$$

where N is the number of detected features; j is the feature index; and k_j can be easily obtained by inverting Equation (13), noting that the attenuated radiance in the left and right images are the same:

$$k_j = -\frac{1}{\|\mathbf{x}_{\text{obj},j}^L\|} \ln \left[1 - \frac{I^L(\mathbf{x}_{\text{obj},j}^L) - I^R(\mathbf{x}_{\text{obj},j}^R)}{S_\infty^L(\mathbf{x}_{\text{obj},j}^L) - S_\infty^R(\mathbf{x}_{\text{obj},j}^R)} \right]. \quad (\text{A2})$$

Figure A1 depicts an estimation of the constant k and the effect of the active polarization on k . Figure A1a shows an example of feature matching between left and right images using SURF. From these matched features, k can be estimated by using Equations (A1) and (A2). The method was applied to images of different attenuation coefficient c and two different polarization angles, namely, 45° and 90° . To change states of the polarization, we fix the polarizers of the two cameras and rotate the polarizer in front of the light source. Figure A1b illustrates the relationship of k and c of two different states of polarization. Solid lines show the value of k when the polarization angle is 45° and dash lines depict the value when the states of polarizers of the light and the cameras are orthogonal. The figure

shows that when c is smaller than 0.8 m^{-1} , k and c are close each other. Additionally, k of values of different polarization angles are almost identical. When c is beyond 0.8 , k in the solid lines increase dramatically with c . The dash lines, however, almost have a linear relationship with c . We could not measure k in denser conditions since very few features can be detected and matched. However, it can be proven that the orthogonal setup of polarization between the illuminator and receivers provides an advantage in scattering reduction because the rate at which $S(\mathbf{x}_{\text{obj}}^i)$ increases with $\|\mathbf{X}_{\text{obj}}^i\|$ set by parameter k . It is worth noting that k also depends on the lighting geometry; thus, the values shown in Figure A1 are only correct in the specific lighting geometry in our experiment. The method to measure c is explained in Section 4.1.

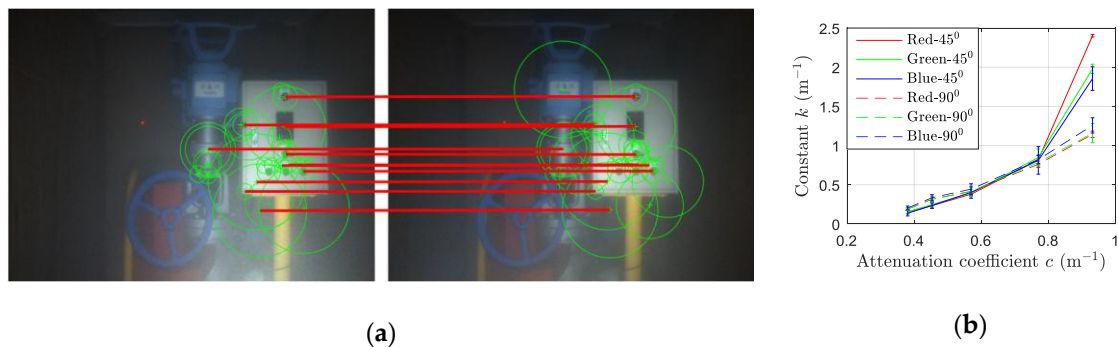


Figure A1. Effect of polarization: (a) Correspondences between left-right images; (b) Constant k estimation, which is attenuation coefficient dependent.

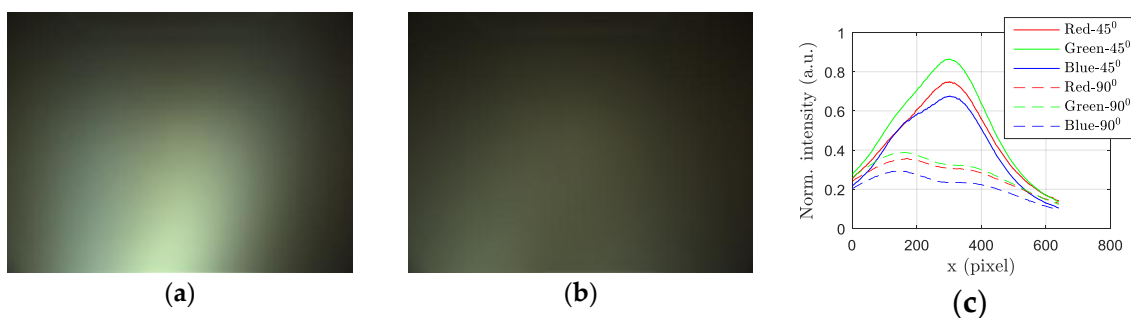


Figure A2. Effect of polarization in saturated backscattering: (a) Saturated backscattering of left camera when the polarization angle is 45° ; (b) Saturated backscattering of left camera when the angle of polarization is 90° ; (c) Intensity along x-axis at the section where intensity attains the highest value.

Figure A2 shows another advantage of using polarization. Besides reducing k , the orthogonal setup can reduce as much as 50% of saturated backscattering of the 45° setup where the signal attains the highest value (Figure A2c). From these two figures above, it is proven that, with slight modification of the system, we can significantly remove the scattering signal. It is also well known that the specular reflection often leads to problems in stereo matching. Therefore, by removing specular reflection, we can also improve stereo matching.

References

1. Starr, J.W.; Lattimer, B.Y. Evaluation of Navigation Sensors in Fire Smoke Environments. *Fire Technol.* **2014**, *50*, 1459–1481. [[CrossRef](#)]
2. Massot-Campos, M.; Oliver-Codina, G. Optical Sensors and Methods for Underwater 3D Reconstruction. *Sensors* **2015**, *15*, 31525–31557. [[CrossRef](#)] [[PubMed](#)]
3. Chi, S.; Xie, Z.; Chen, W. A Laser Line Auto-Scanning System for Underwater 3D Reconstruction. *Sensors* **2016**, *16*, 1534. [[CrossRef](#)] [[PubMed](#)]

4. Narasimhan, S.G.S.G.; Nayar, S.K.S.K.; Bo Sun, B.; Koppal, S.J. Structured light in scattering media. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 15–21 October 2005; pp. 420–427.
5. Bianco, G.; Gallo, A.; Bruno, F.; Muzzupappa, M. A Comparative Analysis between Active and Passive Techniques for Underwater 3D Reconstruction of Close-Range Objects. *Sensors* **2013**, *13*, 11007–11031. [[CrossRef](#)] [[PubMed](#)]
6. Bräuer-Burchardt, C.; Heinze, M.; Schmidt, I.; Kühmstedt, P.; Notni, G. Underwater 3D Surface Measurement Using Fringe Projection Based Scanning Devices. *Sensors* **2016**, *16*, 13. [[CrossRef](#)] [[PubMed](#)]
7. Bodenmann, A.; Thornton, B.; Ura, T. Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light. *J. Field Robot.* **2016**. [[CrossRef](#)]
8. Scharstein, D.; Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [[CrossRef](#)]
9. Scharstein, D.; Hirschmüller, H.; Kitajima, Y.; Krathwohl, G.; Nešić, N.; Wang, X.; Westling, P. High-resolution stereo datasets with subpixel-accurate ground truth. *Lect. Notes Comput. Sci.* **2014**, *8753*, 31–42.
10. Lazaros, N.; Sirakoulis, G.C.; Gasteratos, A. Review of Stereo Vision Algorithms: From Software to Hardware. *Int. J. Optomechatronics* **2008**, *2*, 435–462. [[CrossRef](#)]
11. Tippetts, B.; Lee, D.J.; Lillywhite, K.; Archibald, J. Review of stereo vision algorithms and their suitability for resource-limited systems. *J. Real-Time Image Process.* **2013**, *11*, 5–25. [[CrossRef](#)]
12. Faugeras, O.; Viéville, T.; Theron, E.; Vuillemin, J.; Hotz, B.; Zhang, Z.; Moll, L.; Bertin, P.; Mathieu, H.; Fua, P.; et al. Real-time Correlation-Based Stereo: Algorithm, Implementations and Applications. In *Research Report RR-2013*; INRIA: Versailles, Yvelines, France, 1993.
13. Mühlmann, K.; Maier, D.; Hesser, J.; Männer, R. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *Int. J. Comput. Vis.* **2002**, *47*, 79–88. [[CrossRef](#)]
14. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 650–656. [[CrossRef](#)] [[PubMed](#)]
15. Hosni, A.; Bleyer, M.; Gelautz, M.; Rhemann, C. Local stereo matching using geodesic support weights. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 2093–2096.
16. Hosni, A.; Rhemann, C.; Bleyer, M.; Rother, C.; Gelautz, M. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 504–511. [[CrossRef](#)] [[PubMed](#)]
17. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
18. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [[CrossRef](#)]
19. Kolmogorov, V.; Zabih, R. Computing Visual Correspondence with Occlusions via Graph Cuts. In Proceedings of the Eighth IEEE International Conference on Computer Vision, 2001, (ICCV 2001), Vancouver, BC, Canada, 7–14 July 2001. [[CrossRef](#)]
20. Sun, J.; Zheng, N.N.; Shum, H.Y. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 787–800.
21. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient belief propagation for early vision. *Int. J. Comput. Vis.* **2006**, *70*, 41–54. [[CrossRef](#)]
22. Schechner, Y.Y.; Narasimhan, S.G.; Nayar, S.K. Polarization-based vision through haze. *Appl. Opt.* **2003**, *42*, 511–525. [[CrossRef](#)] [[PubMed](#)]
23. Treibitz, T.; Schechner, Y.Y. Instant 3Descatter. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Stanford, CA, USA, 17–22 June 2006; pp. 1861–1868.
24. Treibitz, T.; Schechner, Y.Y. Active polarization descattering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 385–399. [[CrossRef](#)] [[PubMed](#)]
25. McCartney, E.J. *Optics of the Atmosphere: Scattering by Molecules and Particles*; Wiley: Hoboken, NJ, USA, 1976.
26. Tan, R.T. Visibility in bad weather from a single image. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 23–28 June 2008.
27. Fattal, R. Single Image Dehazing. *ACM Trans. Graphics* **2008**, *27*, 72. [[CrossRef](#)]
28. Nishino, K.; Kratz, L.; Lombardi, S. Bayesian defogging. *Int. J. Comput. Vis.* **2012**, *98*, 263–278. [[CrossRef](#)]

29. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**. [[CrossRef](#)]
30. Negru, M.; Nedevschi, S.; Peter, R.I. Exponential Contrast Restoration in Fog Conditions for Driving Assistance. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2257–2268. [[CrossRef](#)]
31. Chen, C.; Do, M.N.; Wang, J. *Robust Image and Video Dehazing with Visual Artifact Suppression via Gradient Residual Minimization*; Springer: Berlin, Germany, 2016; pp. 576–591.
32. Lee, S.; Yun, S.; Nam, J.-H.; Won, C.S.; Jung, S.-W. A review on dark channel prior based image dehazing algorithms. *EURASIP J. Image Video Process.* **2016**, 2016. [[CrossRef](#)]
33. Zhu, Q.; Mai, J.; Shao, L. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533. [[PubMed](#)]
34. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [[CrossRef](#)]
35. Zhang, J.; Cao, Y.; Wang, Z. Nighttime haze removal based on a new imaging model. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4557–4561.
36. Li, Y.; Tan, R.T.; Brown, M.S. Nighttime Haze Removal with Glow and Multiple Light Colors. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Los Alamitos, CA, USA, 7–13 December 2015; pp. 226–234.
37. Caraffa, L.; Tarel, J.-P. Combining Stereo and Atmospheric Veil Depth Cues for 3D Reconstruction. *IPSP Trans. Comput. Vis. Appl.* **2014**, *6*, 1–11. [[CrossRef](#)]
38. Roser, M.; Dunbabin, M.; Geiger, A. Simultaneous underwater visibility assessment, enhancement and improved stereo. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation, Hong Kong, China, 31 May–7 June 2014.
39. Li, Z.; Tan, P.; Tan, R.T.; Zou, D.; Zhou, S.Z.; Cheong, L.-F. Simultaneous video defogging and stereo reconstruction. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4988–4997.
40. Negahdaripour, S.; Sarafraz, A. Improved stereo matching in scattering media by incorporating a backscatter cue. *IEEE Trans. Image Process.* **2014**, *23*, 5743–5755. [[CrossRef](#)] [[PubMed](#)]
41. Schechner, Y.Y.; Karpel, N. Recovery of Underwater Visibility and Structure by Polarization Analysis. *IEEE J. Ocean. Eng.* **2005**, *30*, 570–587. [[CrossRef](#)]
42. Debevec, P.E.; Malik, J. Recovering high dynamic range radiance maps from photographs. In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97), Los Angeles, CA, USA, 3–8 August 1997; pp. 369–378.
43. Kim, H.; Jin, H.; Hadap, S.; Kweon, I. Specular reflection separation using dark channel prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Oregon, Portland, 25–27 June 2013.
44. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [[CrossRef](#)] [[PubMed](#)]
45. Criminisi, A.; Blake, A.; Rother, C.; Shotton, J.; Torr, P.H.S. Efficient Dense Stereo with Occlusions for New View-Synthesis by Four-State Dynamic Programming. *Int. J. Comput. Vis.* **2007**, *71*, 89–110. [[CrossRef](#)]
46. Giakos, G.C. Active backscattered optical polarimetric imaging of scattered targets. In Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 04CH37510), Como, Italy, 18–20 May 2004; pp. 430–432.
47. Gilbert, G.D.; Pernicka, J.C. Improvement of Underwater Visibility by Reduction of Backscatter with a Circular Polarization Technique. *Appl. Opt.* **1967**, *6*, 741. [[CrossRef](#)] [[PubMed](#)]
48. Lewis, G.D.; Jordan, D.L.; Roberts, P.J. Backscattering target detection in a turbid medium by polarization discrimination. *Appl. Opt.* **1999**, *38*, 3937. [[CrossRef](#)] [[PubMed](#)]
49. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]

