Simultaneous Estimation of Object Region and Depth in Participating Media Using a ToF Camera

Yuki FUJIMURA \(^{(a)}\), Motoharu SONOGASHIRA \(^{(b)}\), Nonmembers, and Masaaki IYAMA \(^{(c)}\), Member

SUMMARY

Three-dimensional (3D) reconstruction and scene depth estimation from 2-dimensional (2D) images are major tasks in computer vision. However, using conventional 3D reconstruction techniques gets challenging in participating media such as murky water, fog, or smoke. We have developed a method that uses a continuous-wave time-of-flight (ToF) camera to estimate an object region and depth in participating media simultaneously. The scattered light observed by the camera is saturated, so it does not depend on the scene depth. In addition, received signals bouncing off distant points are negligible due to light attenuation, and thus the observation of such a point contains only a scattering component. These phenomena enable us to estimate the scattering component in an object region from a background that only contains the scattering component. The problem is formulated as robust estimation where the object region is regarded as outliers, and it enables the simultaneous estimation of an object region and depth on the basis of an iteratively reweighted least squares (IRLS) optimization scheme. We demonstrate the effectiveness of the proposed method using captured images from a ToF camera in real foggy scenes and evaluate the applicability with synthesized data.

key words: time-of-flight, depth estimation, participating media, light scattering, iteratively reweighted least squares

1. Introduction

Three-dimensional (3D) reconstruction and scene depth estimation from 2-dimensional (2D) images are important tasks in computer vision. These techniques can be utilized for a variety of applications such as robot vision and self-driving vehicles. However, using 3D reconstruction techniques gets challenging in participating media such as murky water, fog, or smoke. As shown at the bottom left of Fig. 1 (a), the contrast of a captured image in participating media is reduced because light propagating through the media gets attenuated and the received signal contains not only reflected light from the object surface but also scattered light due to suspended particles. These effects make it difficult to use conventional 3D reconstruction techniques such as structure from motion or shape from shading.

In response to the issue above, we have developed a method that enables us to acquire 3D geometry in participating media. Specifically, we use a time-of-flight (ToF) camera to detect an object and estimate the distance in participating media such as a foggy scene. (b) Overview of proposed method. A continuous-wave ToF camera captures an amplitude image and a phase image. From these images captured in participating media, we estimate the object region and recover depth simultaneously.

camera to detect an object and estimate the distance in participating media (see Fig. 1 (b)). The proposed method enables simultaneous automatic obstacle detection and depth reconstruction in challenging environments such as disaster sites where light is scattered.

There are several architectures for ToF cameras. We use a continuous-wave ToF camera that emits a modulated sinusoid signal into a scene and then measures the amplitude of light that bounces off an object surface and the phase shift between the illumination and received signal. These observations are represented as an amplitude image and a phase image as shown in Fig. 1 (a). Since the phase shift depends on an optical path, we can reconstruct the depth from the phase shift. In this study, we denote the observation of an object surface by “direct component.”

This architecture assumes that each camera pixel ob-
serves a single point in a scene. As mentioned previously, however, the observed signal in participating media includes a scattering component due to light scattering as well as a direct component. The amplitude and phase shift suffer from the scattering effect, and this causes error of depth measurement (see Fig. 1 (a)).

To recover the direct component, we leverage the saturation of a scattering component and light attenuation in participating media. Given a near light source in participating media, a scattering component is saturated close to the camera [1], [2]. This means the scattering component does not depend on an object in the scene if the object is located at a certain distance from the camera. Moreover, the intensity of light propagating through participating media is attenuated exponentially relative to distance. Thus, reflected light from a distant point is negligible and the observation of such a point includes only a scattering component. In this paper, we assume that a target scene consists of an object region and a background that only contains a scattering component. This scattering component can then be estimated simply by observing the background. In addition to the above assumption, we introduce two priors to estimate the scattering component: first, a scattering component can be approximated using a quadratic function in a local image patch (local quadratic prior), and second, a scattering component has a symmetrical characteristic in an overall image (global symmetrical prior).

The estimation of a scattering component is formulated as a robust estimation, where the object region is regarded as outliers because it contains a direct component. We propose an optimization scheme based on iteratively reweighted least squares (IRLS) [3]–[6], which minimizes weighted least squares iteratively. The object region can then be extracted via the IRLS weight as outliers.

The primary contributions of this paper are summarized as follows: (1) we leverage the saturation and attenuation of the scattering and direct component to improve a depth measurement of a continuous wave ToF camera in participating media; (2) the formulation of robust estimation and IRLS optimization scheme with the proposed priors enable the effective estimation of the scattering component and object region at the same time.

2. Related Work

2.1 Scattering Removal

In participating media, the contrast of a captured image is reduced due to scattered light and light attenuation. In computer vision and image processing, several methods have been proposed to recover an original image from a degraded one [7]–[10]. These methods are referred to as defogging or dehazing, and the problem is known to be ill-posed. Many priors have been introduced to address the ill-posed nature of the problem. For example, He et al. [7] and Berman et al. [10] respectively proposed a dark channel prior and a haze-line prior. Recently, many learning-based methods using neural networks have also been proposed [11]–[15].

Almost all of these methods are based on an atmospheric scattering model under daylight [16], which is the simple combination of the direct component and the scattering component. However, the model does not hold under active light source such as a ToF camera. Instead, we model the observation using a single scattering model (Sect. 3). Our problem setting is similar to the scattering removal of a RGB image under a near light source [2]. We take an amplitude and a phase image captured by a ToF camera as input for scattering removal with the proposed priors (Sect. 4.1).

2.2 3D Reconstruction in Participating Media

Our goal is to reconstruct a 3D scene in participating media. There have been several works on the design of 3D reconstruction techniques for participating media. For example, Li et al. [17] proposed simultaneous dehazing and 3D reconstruction using multi-view stereo. Photometric stereo in participating media has been proposed by [2], [18], [19], in which light transport models were built in participating media for photometric stereo application. Other studies have been based on structured light [20], [21], light field [22], and the absorption of infrared light in underwater scenes [23].

2.3 Multipath Interference of ToF

A ToF camera assumes that each camera pixel observes a single point in a scene. In participating media, however, the measurement also includes scattered light. This problem is known as multipath interference (MPI). MPI is caused not just by light scattering in participating media but also by subsurface scattering or interreflection in common scenes. Thus, many previous studies have tackled MPI compensation [24]–[28].

In this paper, we limit our focus to MPI caused by light scattering in participating media. ToF measurement in participating media has been proposed by [29], [30]. Heide et al. [29] developed a scattering model based on exponentially modified Gaussians for transient imaging using a photonic mixer device (PMD) [31]. Satat et al. [30] demonstrated that scattered photons observed with a single photon avalanche diode (SPAD) have gamma distribution and leveraged this observation to separate received photons into a directly reflected component and a scattering component. Our method differs from these approaches in that we just use an off-the-shelf ToF camera such as Microsoft Kinect v2 with no special hardware modification. The concurrent work by Muraji et al. [32] also used a continuous-wave ToF camera. They removed scattering effect using multiple modulation frequencies. We address different problem settings as follows: (1) we model spatially varying scattering components due to a limited lighting angle as explained in Sect. 4.1; (2) we model the simultaneous estimation of object regions and scattering components as a single optimization problem.
3. ToF Observation in Participating Media

In this section, we describe our image formation model for a continuous-wave ToF camera in participating media. As with many previous studies[11], [2], [20], [33], we assume here that forward scattering and multiple scattering are negligible and that the density of a participating medium in a scene is homogeneous.

A continuous-wave ToF camera illuminates a scene with amplitude-modulated light and then measures the amplitude of received signal $\alpha$ and phase shift $\varphi$ between the illumination and received signal. This observation can be described using a phasor [34], as

$$\alpha e^{j\varphi} \in \mathbb{C},$$

(1)

Since the phase shift is proportional to the depth of an object, we can compute the depth as

$$z = \frac{c \varphi}{4\pi f},$$

(2)

where $z$ is depth, $c$ is the speed of light, and $f$ is the modulation frequency of the camera.

In participating media, the observation contains scattered light. As shown in Fig. 2, light interacts with the medium on the line of sight and then arrives at the camera pixel. Thus, the observed scattering component is the sum of scattered light on the line of sight. Now, we consider the 3D coordinate, the origin of which is the camera center. When the camera observes a surface point $p^* \in \mathbb{R}^3$ at a camera pixel $(u, v)$, the total observation $\tilde{\alpha}(u, v; p^*)e^{j\varphi(u, v; p^*)}$ can be written as

$$\tilde{\alpha}(u, v; p^*)e^{j\varphi(u, v; p^*)} = \alpha_d(u, v; p^*)e^{j\varphi_d(u, v; p^*)} + \int_{\|p\| \geq \|p_{saturate}\|} \alpha_s(u, v; p)e^{j\varphi_s(u, v; p)} d\|p\|,$$

(3)

where $\alpha_d(u, v; p^*)$ and $\varphi_d(u, v; p^*)$ are the direct components, $\alpha_s(u, v; p)$ and $\varphi_s(u, v; p)$ are the scattering components, which depend on only the camera pixel $(u, v)$ rather than the object depth.

Although the observation consists of the direct component $\alpha_d(u, v; p^*)e^{j\varphi_d(u, v; p^*)}$ and the scattering component $\alpha_s(u, v)e^{j\varphi_s(u, v)}$, the attenuation due to the medium reduces the direct component dramatically. Thus, if the camera observes a distant point $p_{far}$, the amplitude of the reflected light fades away, that is,

$$\alpha_d(u, v; p_{far}) = 0.$$

(6)

Therefore, the observation of the distant point includes only a scattering component:

$$\tilde{\alpha}(u, v; p_{far})e^{j\varphi(u, v; p_{far})} = \alpha_s(u, v)e^{j\varphi_s(u, v)}.$$

(7)

Figure 3 shows amplitude and phase images when the camera observes a black surface in a foggy scene. The intensity of reflected light from the black surface is very small, so this approximates a distant observation where only a scattering component can be observed. As discussed above, in both the amplitude and phase images, the scattering component is inhomogeneous because the illumination has a limited beam

![Fig. 2 Light scattering in participating media. Light interacts with a participating medium on the line of sight and then arrives at a camera pixel. The total scattering component is the sum of scattered light on the red line in the figure, which depends on the limited beam angle of the light source.](Image)

![Fig. 3 Observation of a black surface in a foggy scene. The black surface approximates a distant observation where only a scattering component can be observed because reflected light from the scene gets attenuated. Note that the observed scattering component is inhomogeneous due to the limited beam angle of the illumination.](Image)
angle.

The proposed method is based on the assumption that the scattering component is saturated close to the camera. However, this assumption is not precise in some cases, for example, when a scene is extremely close to the camera. The measurement range of our method is between a saturation point and a background point that has no direct component. We investigate the effective range of the proposed method in Sect. 5.3.

4. Simultaneous Estimation of Object Region and Depth

As explained in the previous section, a scattering component depends on the position of a camera pixel rather than a target object. In addition, only the scattering component is observed in the background where an object is farther away. Thus, our goal is to estimate the scattering component in an object region from a background because the component does not depend on the position of a camera pixel rather than a background point that has no direct component. We investigate the effective range of the proposed method in Sect. 5.3.

4.1 Prior of Scattering Component

We can estimate the scattering component of an object region from a background because the component does not depend on the object. Tsiotsios et al. [2] approximated backscatter as a quadratic function in a captured image. Similarly to their work, we also introduce priors, local quadratic prior and global symmetrical prior, that allow us to estimate the scattering component.

(1) Local quadratic prior

In our ToF setting, we found that a scattering component cannot be approximated globally with a simple function such as Tsiotsios et al. [2] where a scattering component is fitted with a quadratic function all over an image. Thus, as shown in Fig. 4, we assume that a scattering component can be represented with a quadratic function in a local image patch, that is,

\[ x_k(u, v) = a_1^k u^2 + a_2^k uv + a_3^k v^2 + a_4^k u + a_5^k v + a_6^k \]

where \( x_k(u, v) \) is the value at a pixel \((u, v)\) in a local image patch \( x_k \). \( u = [u^2 \ uv \ v^2 \ u \ v \ v-1] \) is a 6-dimensional vector and \( \mathbf{a}_k = [a_1^k, a_2^k, a_3^k, a_4^k, a_5^k, a_6^k] \) denotes the coefficients of the quadratic function in patch \( x_k \).

(2) Global symmetrical prior

However, this local prior is not useful when there exists a large object region and a quadratic function is also fitted into the values in that region. To address this problem, we introduce a global prior to the scattering component.

As discussed in Sect. 3, a scattering component depends on the relative position between a camera pixel and a light source. This is because the individual starting points of the integral in Eq. (3) differ from each other. Meanwhile, as shown in Fig. 5, we assume that the camera and light source are collocated on a line that is parallel to the horizontal axis of the image, the observed scattering component has symmetry because the integral domain of a pixel is consistent with that of the symmetrical pixel with respect to the central axis of the image.

As explained in the previous section, a scattering component is observed in the background where an object is farther away. However, this assumption is not precise in some cases, for example, when a scene is extremely close to the camera. The measurement range of our method is between a saturation point and a background point that has no direct component. We investigate the effective range of the proposed method in Sect. 5.3.

In this section, we describe how our method divides camera pixels into an object region and a background, and simultaneously estimates the scattering component in the object region. First, we introduce two priors to estimate the scattering component, and then the problem is formulated as robust estimation, which allows us to extract the object region as outliers. In the following, with a slight alteration of notation, we refer to both an amplitude image and a phase image as an image, since we process both images in the same manner.

In our ToF setting, we found that a scattering component cannot be approximated globally with a simple function such as Tsiotsios et al. [2] where a scattering component is fitted with a quadratic function all over an image. Thus, as shown in Fig. 4, we assume that a scattering component can be represented with a quadratic function in a local image patch, that is,

\[ x_k(u, v) = a_1^k u^2 + a_2^k uv + a_3^k v^2 + a_4^k u + a_5^k v + a_6^k \]

where \( x_k(u, v) \) is the value at a pixel \((u, v)\) in a local image patch \( x_k \). \( u = [u^2 \ uv \ v^2 \ u \ v \ v-1] \) is a 6-dimensional vector and \( \mathbf{a}_k = [a_1^k, a_2^k, a_3^k, a_4^k, a_5^k, a_6^k] \) denotes the coefficients of the quadratic function in patch \( x_k \).

(2) Global symmetrical prior

However, this local prior is not useful when there exists a large object region and a quadratic function is also fitted into the values in that region. To address this problem, we introduce a global prior to the scattering component.

As discussed in Sect. 3, a scattering component depends on the relative position between a camera pixel and a light source. This is because the individual starting points of the integral in Eq. (3) differ from each other. Meanwhile, as shown in Fig. 5, we assume that the camera and light source are collocated on the line that is parallel to the horizontal axis of the image. ToF devices can easily be built on the basis of this setting (e.g., Microsoft Kinect v2 has this setting). In this case, the integral domain of a pixel is consistent with
that of the symmetrical pixel with respect to the central axis of the image. Thus, the observed scattering component also has symmetry, and we leverage this symmetry as a global prior. In the appendix, we describe the symmetrical prior mathematically.

4.2 Formulation as Robust Estimation

We formulate the scattering component estimation problem as robust estimation. Specifically, we solve the following optimization problem:

\[
\min_{x,a_1,\ldots,a_K} \sum_{i=1}^{N} \rho \left( \frac{x_i - \tilde{x}_i}{\sigma_1} \right) + \gamma_1 \sum_{k=1}^{K} \|Ua_k - x_k\|^2 + \gamma_2\|Fx - x\|^2 + \gamma_3\|\nabla x\|^2.
\]  

(11)

The first term of Eq. (11) is a data term where \( \tilde{x} = [\tilde{x}_1 \cdots \tilde{x}_N]^T \) and \( x = [x_1 \cdots x_N]^T \) are a captured image and a scattering component, respectively. \( N \) is the number of camera pixels, and \( \sigma_1 \) is a scale parameter. We use a non-linear differentiable function \( \rho(x) \) rather than square error \( x^2 \), which allows us to make the estimation robust against outliers. In this study, we simply use the residual of the observation and the scattering component as the data term, i.e., pixels that contain a direct component are regarded as outliers.

We use three additional regularization terms. The second term represents the local prior. \( K \) is the number of patches for local quadratic function fitting. \( U \) is an \( N_k \times 6 \) matrix where \( N_k \) is the number of pixels in patch \( x_k \in \mathbb{R}^{N_k} \) and each row of \( U \) is a vector \( u \) that corresponds to each pixel coordinate. In this study, these patches do not overlap each other. The third term represents the global prior where \( F \in \mathbb{R}^{N \times N} \) is a matrix that flips an image vertically. The last term is a smoothing term where \( \nabla \) denotes a gradient operator. This smoothing accelerates the optimization. Hyperparameters \( \gamma_1, \gamma_2, \gamma_3 \) control the contribution of each term.

4.3 IRLS and Object Region Estimation

We minimize Eq. (11) with respect to a scattering component \( x \) and the coefficients of quadratic functions \( a_1, \ldots, a_K \). However, the nonlinearity of \( \rho(x) \) makes it difficult to obtain a closed-form solution. For efficient computation, the IRLS optimization was developed in the literature [3], [4]. IRLS minimizes weighted least squares iteratively and the weight is updated using the current estimate in each iteration. The objective function in Eq. (11) is transformed into weighted least squares as follows:

\[
\min_{x,a_1,\ldots,a_K} (x - \tilde{x})^T W (x - \tilde{x}) + \gamma_1 \sum_{k=1}^{K} \|Ua_k - x_k\|^2 + \gamma_2\|Fx - x\|^2 + \gamma_3\|\nabla x\|^2.
\]  

(12)

where \( W = \text{diag}(w) \) is an \( N \times N \) matrix and \( w = [w_1, \ldots, w_N]^T \) is the weight for each error \( x_i - \tilde{x}_i \). Hyperparameters are given as \( \gamma = 2\sigma_1^2 \gamma_0 \). Equation (12) is quadratic with respect to the scattering component \( x \), and thus is easy to optimize. In each iteration, we solve Eq. (12) for \( x \) and \( a_1, \ldots, a_K \), and the weight can be updated using the current estimate as

\[
w_i = \frac{\rho'(x_i - \tilde{x}_i) / \sigma_1}{(x_i - \tilde{x}_i) / \sigma_1}.
\]  

(13)

The specific update rule of the weight depends on the nonlinear function \( \rho(x) \). In this study, we use the following function as \( \rho(x) \):

\[
\rho(x) = \begin{cases} 
\frac{c}{2} & \text{if } |x| \leq c \\
\frac{c}{2} \left(1 - \left(\frac{x}{c}\right)^2\right)^3 & \text{if } |x| > c
\end{cases}
\]  

(14)

This function yields the following update:

\[
w_i = \begin{cases} 
1 - \left(\frac{r_i}{c}\right)^2 & \text{if } |r_i| \leq c \\
0 & \text{otherwise}
\end{cases}
\]  

(15)

where \( r_i = (x_i - \tilde{x}_i) / \sigma_1 \), and \( c \) is a tuning parameter. This update is referred to as Tukey’s biweight [4], [35], where \( 0 \leq w_i \leq 1 \).

(1) Object region from IRLS weight

The weight controls the robust estimation, that is, a large error term reduces the corresponding weight. In this study, we consider the object region as outliers, and thus the weight in the object region should be small. Therefore, we can leverage the IRLS weight to extract the object region from the image.

(2) Selection of \( \rho(x) \)

In Eq. (11), any robust functions can be selected for \( \rho(x) \) (e.g., Huber loss [36] or t-distribution [37]). However, we found that Tukey’s biweight [4], [35] is suitable because of its property. The robustness of Tukey’s biweight is achieved by truncating outliers (Eq. (15)), which is an object region in our case. This property results in a clear object boundary as shown in the experiments. In contrast, the weight of Huber loss, for example, has long tails, and this makes the optimization unstable.

4.4 Coarse-to-Fine Optimization

The accurate object region extraction is critical for the effectiveness of the scattering component estimation. In Sect. 4.1, we introduced the local and global priors of the scattering component to deal with a large object region. To make the region extraction more robust, we developed a coarse-to-fine optimization scheme. Before solving Eq. (11), we optimize the following objective function:

\[
\min_{x,a_1,\ldots,a_K} \sum_{k=1}^{K} \rho \left( \frac{\|x_k - \tilde{x}_k\|}{\sigma_2} \right) + \gamma_1 \sum_{k=1}^{K} \|Ua_k - x_k\|^2
\]  

(16)
Algorithm 1 Simultaneous estimation of scattering component and object region

Require: Image $\tilde{x}$
Ensure: Scattering component $x$ and object mask $w$

Coarse level optimization (Eq. (16)):

$W \leftarrow I$, $a_k \leftarrow \arg\min_{a_k} \|Ua_k - \tilde{x}_k\|^2$

repeat
  Solve Eq. (12) for $x$
  Solve Eq. (12) for $a_1, \cdots, a_K$
  if first iteration then
    Compute $\sigma_2$
  end if
  Update $w$ in patch-wise
until converged

Fine level optimization (Eq. (11)):

Initialize $w$ and $a_1, \cdots, a_K$ with the output of the coarse level

repeat
  Solve Eq. (12) for $x$
  Solve Eq. (12) for $a_1, \cdots, a_K$
  if first iteration then
    Compute $\sigma_1$
  end if
  Update $w$ in pixel-wise
until converged

Binarize $w$

$$+ \gamma_2 \|Fx - x\|^2 + \gamma_3 \|\nabla x\|^2.$$ \hspace{1cm} (16)

This is similar to the patch-based robust regression proposed by [38]. The difference from Eq. (11) is that the data term consists of patch-wise errors. Equation (16) can be transformed into IRLS as well as Eq. (11) where $\gamma_i = 2\sigma_i^2 \gamma_{\ast}$, and the IRLS weight is updated patch-wise rather than pixel-wise.

Algorithm 1 shows the overall procedure of the simultaneous estimation of a scattering component and an object region. We first solve Eq. (16) for the weight in a patch level and then solve (11) in a pixel level. Each scale parameter is computed only at the first iteration and is fixed during subsequent iterations. We compute the scale parameters using a median absolute deviation, which is the robust measure of a deviation [4]. At the end of the algorithm, we binarize the IRLS weight to generate an object mask. This procedure is applied to an amplitude and a phase image in the same manner, and thus we can obtain the object mask in each domain. In this study, we determine the final object mask as their intersection.

5. Experiments

We evaluated the effectiveness of the proposed method using real and synthetic data. First, we show the experiments with real data, and then, the applicability to various scenes is discussed using synthetic data. Finally, we investigate the effective measurement range of the proposed method. All experiments were done on Intel Core i5@3.1GHz with 8GB RAM and code was written in Python.

5.1 Experiments with Real Data

First, we performed the experiments in a scene shown in Fig. 6. We set up a fog generator and a Kinect v2 in a closed space sized 186 \times 161 cm with black walls and floor. This ToF camera has an IR camera and a light source inside it. This is a common configuration in ToF cameras. The light source emits modulated infrared light. Fog generated by the fog machine is made of liquid of propylene glycol. After generating fog, we waited for a few minutes so that fog is filled in the closed space and regarded as a homogeneous medium. The observation of the wall includes only a scattering component because incident light into the wall is absorbed. The Kinect v2 has three modulation frequencies: 120, 80, and 16 MHz. We used images obtained with 16 MHz. Larger frequencies have shorter depth measurement range due to phase wrapping. The measurement capability of 16 MHz is about 9 meters, which is the largest measurement range of the frequencies of the Kinect v2. To acquire an amplitude and phase image, we used the source code given by [39]. Their code provides the average image of several frames, and we modified the code so that only a single frame was input. If we use multiple frames, the estimation will get to be more robust, while our method works well given only a single frame as shown in the experiments. To compensate for high frequency noise, we used a bilateral filter as preprocessing.

The spatial resolution of an image captured by Kinect v2 is 424 \times 512 pixels. We divided a captured image into 4 \times 4 patches ($K = 16$) for local quadratic prior. In Sect. 4.1, we assumed that the camera and the light source are collocated on a line that runs parallel to the horizontal axis of the image. In practice, the camera and light source in the Kinect v2 are slightly out of alignment, and it is difficult to modify the setup. Although this violates the symmetry of the scattering component, we found that error due to this misalignment is negligibly small as shown in Figs. 3 (b) (c). In our implementation, we defined $F$ as a matrix that flips an image with respect to the 200th row of the image. In addition, we did not use the 24 rows of the lower part of the image for the third term of Eq. (11), as these pixels have no information of global symmetrical prior. Just modifying the flip matrix $F$ by changing the flip center is enough to estimate scattering components instead of modifying the hardware setup. For amplitude images, we set the hyperparameters of the objective function as $[\gamma_1', \gamma_2', \gamma_3'] = [0.1, 0.1, 10]$, and the tuning...
The parameter of the function \( \rho(x) \) is set as \( c = 4, 7 \) in the coarse and fine level optimization, respectively. For phase images, we set \( [\gamma_1', \gamma_2', \gamma_3'] = [0.01, 0.1, 50] \) and \( c = 2, 3 \). The numbers of IRLS iterations were 5 and 50 for the coarse and fine optimization. One iteration required about from 0.3 to 1.0 seconds.

The results are shown in Figs. 7, 8, and 9. We tested the proposed method under different density conditions for investigating the robustness to the density (In Figs. 7, 8, and 9, the density is thin, medium, and thick, respectively).
Table 1  Mean/max depth error on each object of without considering scattering (top) and proposed (bottom) under different density conditions. [cm]

|     | Plane  | Chair  | Desk   | Hand   | Duck   |
|-----|--------|--------|--------|--------|--------|
| Thin| 11.7 / 64.4 | 19.9 / 83.1 | 46.0 / 93.0 | 42.5 / 93.8 | 57.4 / 105.0 |
|     | 1.8 / 47.7  | 6.5 / 89.4  | 8.3 / 70.0  | 3.2 / 43.6  | 4.5 / 39.9  |
| Medium| 25.3 / 97.3 | 37.2 / 98.2 | 65.6 / 92.1 | 67.9 / 102.5 | 79.8 / 111.0 |
|     | 2.1 / 28.5  | 6.0 / 49.8  | 10.7 / 78.6 | 5.1 / 57.1  | 11.8 / 74.9  |
| Thick| 42.1 / 105.5 | 53.1 / 110.6 | 79.9 / 110.4 | 84.4 / 123.2 | 95.4 / 121.9 |
|     | 2.0 / 39.7  | 7.1 / 71.2  | 20.2 / 99.2 | 7.6 / 75.9  | 14.3 / 67.9  |

Fig. 11  Error histograms of each object. (a)–(c) show the error under thin, medium, and thick fog, respectively. The blue and orange bars represent the error without and with the proposed method.

Next, we tested the proposed method in a scene shown in Fig. 12. We artificially generated fog in the same manner as the previous experiments, while the scene in Fig. 12 has neither dark walls nor floor. Note that the scene has materials with various types of reflectance, including a lamp made from paper, a glossy vase, and a wooden ornament. The
Fig. 12 Results of other real scene. (a) Target scene without and with fog. (b)(c) Left to right: input image, estimated scattering component, and IRLS weight for the amplitude and phase image, respectively. (d) Left to right: depth without fog, depth with fog, reconstructed depth, masked reconstructed depth, and estimated object mask.

Fig. 13 Procedure of synthesizing images. First, we captured a scene that has calibration objects in a foggy scene and masked the region of the calibration objects manually. After that, we compensated for the defective region to estimate the scattering component. Using the observation without fog, the scattering coefficient can be computed. The images of a target scene without fog were captured separately, and the attenuated direct component and the estimated scattering component were combined into synthesized images.

estimation of the scattering component and object region for the amplitude and phase image is shown in Fig. 12(b)(c), respectively, and the result of the depth reconstruction is shown in Fig. 12(d). The proposed method could also extract the object region and improve the depth measurement in a scene that has a general background.

5.2 Experiments with Synthesized Data

To investigate the effectiveness in more varied scenes, we evaluated the proposed method with synthesized data. The procedure of generating the synthesized data is shown in Fig. 13. We assume that a scattering component does not depend on object depth, and thus we observed a direct component and a scattering component separately and then combined them into a synthesized image. First, we captured a foggy scene that includes calibration objects for the estimation of the scattering coefficient, and the region of the calibration objects was masked manually. After that, we compensated for the defective region by solving Eq. (12) to estimate the scattering component. In the scene in Fig. 13, a scattering coefficient was computed as $\beta = 3.5 \times 10^{-4}$/mm. We also observed a scene without fog, which was used for the direct component after being attenuated by the scattering coefficient. We combined the attenuated signal and the scattering component to synthesize amplitude and phase images.

The results are shown in Figs. 14, 15, and 16. In each scene, we show (a) the target scene, (b) (c) the estimated scattering component and the IRLS weight for the amplitude and phase image, and (d) the result of the depth reconstruction. In Figs. (a)–(c), the proposed method effectively
extracted the object region and estimated the scattering component. However, in Fig. 15, the center of the object region was regarded as the background. This is due to the quadratic function was partially fitted to the object region. This might be solved by introducing the spatial smoothness of the IRLS weight. We also show a failure case in Fig. 17. In a scene that has a large object region, our method was less effective because a quadratic function also fits to values in the object region. In Fig. 17, a large textureless object region exists on the left side. In addition, the global symmetrical prior

Fig. 14  Results of synthesized data. (a) Target scene. (b)(c) Left to right: input image, estimated scattering component, and IRLS weight for the amplitude and phase image, respectively. (d) Left to right: depth without fog, depth with fog, reconstructed depth, masked reconstructed depth, and estimated object mask.

Fig. 15  Results of synthesized data. (a) Target scene. (b)(c) Left to right: input image, estimated scattering component, and IRLS weight for the amplitude and phase image, respectively. (d) Left to right: depth without fog, depth with fog, reconstructed depth, masked reconstructed depth, and estimated object mask.

Fig. 16  Results of synthesized data. (a) Target scene. (b)(c) Left to right: input image, estimated scattering component, and IRLS weight for the amplitude and phase image, respectively. (d) Left to right: depth without fog, depth with fog, reconstructed depth, masked reconstructed depth, and estimated object mask.
between a background point denoted as $g$. In this section, we discuss the measurement capability of our method in this context. We assume that a scattering component is saturated close to a camera and there exists a background that has only a direct component remains. As shown in Fig. 18, a scene has a saturation point and a background point denoted as $p_{\text{saturate}}$ and $p_{\text{background}}$. For simplicity, the camera and light source are assumed to be collocated in the same place. Far from $p_{\text{saturate}}$, a scattering component is constant due to its saturation, while a direct component fades away far from $p_{\text{background}}$. Therefore, the measurement range of our method is between $p_{\text{saturate}}$ and $p_{\text{background}}$.

We simulated the measurement range to evaluate the capability. Similarly to the process of synthesizing images, a scattering coefficient was computed for the scene in Fig. 8 to use for the simulation ($\beta = 3.2 \times 10^{-4} / \text{mm}$). We use the following Henyey-Greenstein phase function for scattering property:

$$P(\theta) = \frac{1}{4\pi} \left(1 - g^2 \right)^{\frac{1}{2}},$$

where $\theta$ is a scattering angle. The parameter $g$ was set as 0.9 for fog [40]. A scattering component from a camera to depth $z$ is given as

$$\alpha_s(z)e^{i\varphi_s(z)} = \int_{z_0}^z \frac{1}{z} \beta P(\pi) e^{-2\beta z} e^{i\theta} dz.$$ (18)

We set the starting point of the integral as $z_0 = 10$ mm. A direct component from depth $z$ is computed as

$$\alpha_d(z)e^{i\varphi_d(z)} = \frac{L}{z^2} e^{-2\beta z} e^{i\varphi_d},$$ (19)

where $L$ consists of a surface albedo and shading, and we set $L = 1$ in this simulation. The total observation $\tilde{\alpha}(z)e^{i\varphi(z)}$ is the sum of these components as with Eq. (3).

Figure 19 shows the simulation results. In (a) and (b), the horizontal axis denotes depth $z$ and the vertical axis denotes a scattering component for amplitude and phase. These figures validate the saturation characteristic. Meanwhile, in Fig. 19 (c) and (d), the vertical axes denote the residual of the observation and scattering component. $\Delta\varphi$ is given by the residual angle of $\tilde{\alpha}(z)e^{i\varphi(z)}$ and $\alpha_s(z)e^{i\varphi_s(z)}$ on the complex plane. These values represent the remaining direct components from depth $z$.

Now, we can set $||p_{\text{background}}|| = 2500$ mm and $5000$ mm for amplitude and phase from Fig. 19 (c) and (d) because the direct component is close to zero. In contrast, in Fig. 19 (a) and (b), if we set $||p_{\text{saturate}}|| = 1000$ mm, the estimation error of the scattering component for amplitude due to the saturation assumption can be considered almost zero because $1 - \alpha_s(1000)/\alpha_s(8000) \approx 0$, and for phase, the error is $1 - \varphi_s(1000)/\varphi_s(8000) \approx 6.0\%$.

In the experiments with real data, all of the target objects were located between 1000 mm and 2000 mm. If we assume the measurement range between $||p_{\text{saturate}}|| = 1000$ mm and $||p_{\text{background}}|| = 2500$ mm, the target objects are located in that range, and from above discussion, we have just $6.0\%$ error of the scattering component estimation.
6. Conclusion

In this paper, we proposed a method that simultaneously estimates an object region and depth by using a continuous-wave ToF camera. We leveraged the saturation of a scattering component and the attenuation of a direct component from a distant point in a scene. The formulation with a robust estimator and the IRLS optimization scheme allow us to estimate the scattering component and object region simultaneously.

The limitation of the proposed method is that we assume the density of the participating medium in the scene to be homogeneous; thus, it cannot be applied to an inhomogeneous medium or a dynamic floating medium. In these environments, global symmetrical prior does not hold. In addition, we assume that a scene has a background region, which makes it difficult to apply the method to scenes filled with objects. In future work, we will address these problems in order to further enhance the real-world applicability of the proposed method.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 18H03263.

References

[1] T. Treibitz and Y.Y. Schechner, “Active polarization descattering,” IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.3, pp.385–399, 2009.
[2] C. Tsiotsios, M.E. Angelopoulou, T. Kim, and A.J. Davison, “Backscatter compensated photometric stereo with 3 sources,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2259–2266, 2014.
[3] P. Holland and R.E. Welsch, “Robust regression using iteratively reweighted least-squares,” Communications in Statistics A: Theory and Method, vol.6, no.9, pp.813–827, 1977.
[4] J. Fox and S. Weisberg, “Robust regression in R: An appendix to an R companion to applied regression,” SAGE Publishing, 2002.
[5] R. Chartrand and W. Yin, “Iteratively reweighted algorithms for compressive sensing,” The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.3869–3872, 2008.
[6] D. Wipf and S. Nagarajan, “Iterative reweighted 11 and 12 methods for finding sparse solutions,” IEEE J. Sel. Topics Signal Process., vol.3, no.2, pp.317–329, 2010.
[7] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.12, pp.2341–2353, 2011.
[8] K. Nishino, L. Kratz, and S. Lombardi, “Bayesian defogging,” International Journal of Computer Vision, vol.98, no.3, pp.263–278, 2012.
[9] R. Fattal, “Dehazing using color-lines,” ACM Transactions on Graphics (TOG), vol.34, no.1, 2014.
[10] D. Berman, T. Treibitz, and S. Avidan, “Non-local image dehazing,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1674–1682, 2016.
[11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end system for single image haze removal,” IEEE Trans. Image Process., vol.25, no.11, pp.5187–5198, 2016.
In this appendix, we describe the details of the symmetrical prior as a mathematical form. First of all, let symmetrical pixels with respect to the central axis of the image be \((u, v)\) and \((u, \tilde{v})\). The scattering components of these pixels can be given as

\[
\alpha_s(u, v, p) = \int_{\|p\|} \alpha(u, v, p)e^{-j\phi(u, v, p)}d\|p\|, \quad (A.1)
\]

\[
\alpha_s(u, \tilde{v}, p) = \int_{\|p\|} \alpha(u, v, p)e^{-j\phi(u, v, p)}d\|p\|. \quad (A.2)
\]

Then, we substitute \(p\) by

\[
p \leftarrow \hat{p}(u, v), \quad (A.3)
\]

\[
p \leftarrow \hat{p}(u, \tilde{v}), \quad (A.4)
\]

where \(\hat{p}(u, *)\) is the vector of \(p(u, *)\) whose third element is one. \(z\) is the depth of the 3D point \(p\). This substitution yields

\[
\alpha_s(u, v, z) = \int_{z_0(u,v)} \alpha(u, v, z)e^{-j\phi(u, v, z)}d|\hat{p}(u, v)|dz, \quad (A.5)
\]

\[
\alpha_s(u, \tilde{v}, z) = \int_{z_0(u,\tilde{v})} \alpha(u, v, z)e^{-j\phi(u, v, z)}d|\hat{p}(u, \tilde{v})|dz, \quad (A.6)
\]

where \(z_0(u, *) = |\|p_0(u, *)\||/|\hat{p}(u, *)|\) and \(z_{saturate}\) is the depth of the saturated point \(p_{saturate}\). Here, let the position of a light source be \(l \in \mathbb{R}^3\), which is parallel to the horizontal axis. \(\alpha(u, v, z)\) and \(\alpha(u, \tilde{v}, z)\) are given as
\[ \alpha(u, v; z) = \frac{I\beta P(\theta_0)e^{-\beta |z| ||\hat{p}(u, v)|| + ||z - \hat{p}(u, v)||^2}}{||l - \hat{p}(u, v)||^2}, \]  
(A-7)

\[ \alpha(u, \tilde{v}; z) = \frac{I\beta P(\theta_0)e^{-\beta |z| ||\hat{p}(u, \tilde{v})|| + ||z - \hat{p}(u, \tilde{v})||^2}}{||l - \hat{p}(u, \tilde{v})||^2}, \]  
(A-8)

where \( I \) and \( \beta \) are light intensity and a scattering coefficient, respectively. \( P(\theta) \) is a phase function (e.g., Eq. (17)), and the scattering angle \( \theta_1 \) and \( \theta_2 \) are given by

\[ \theta_1 = \arccos \left( \frac{(1 - \hat{p}(u, v))^\top (z - \hat{p}(u, v))}{||l - \hat{p}(u, v)|| ||z - \hat{p}(u, v)||} \right), \]  
(A-9)

\[ \theta_2 = \arccos \left( \frac{(1 - \hat{p}(u, \tilde{v}))^\top (z - \hat{p}(u, \tilde{v}))}{||l - \hat{p}(u, \tilde{v})|| ||z - \hat{p}(u, \tilde{v})||} \right), \]  
(A-10)

Here, from the symmetric property, the following relations hold:

\[ z_0(u, v) = z_0(u, \tilde{v}), \]  
(A-11)

\[ ||\hat{p}(u, v)|| = ||\hat{p}(u, \tilde{v})||, \]  
(A-12)

\[ \Gamma^\top \hat{p}(u, v) = \Gamma^\top \hat{p}(u, \tilde{v}), \]  
(A-13)

\[ ||l - \hat{p}(u, v)|| = ||l - \hat{p}(u, \tilde{v})||. \]  
(A-14)

From Eq. (A-7) to (A-14), we can obtain \( \alpha(u, v; z) = \alpha(u, \tilde{v}; z) \), and this yields \( \alpha_3(u, v; z) = \alpha_3(u, \tilde{v}; z) \) from Eqs. (A-5) and (A-6). Therefore, the scattering component is equal to that of the symmetrical pixel with respect to the horizontal axis.

---

**Yuki Fujimura** received the B.S. degree in engineering and the M.S. degree in informatics from Kyoto University in 2016 and 2018, respectively. He is currently pursuing his Ph.D. degree in informatics at Kyoto University. He is a research fellow (DC2) of Japan Society for the Promotion of Science (JSPS). His research interests include computer vision, 3-D reconstruction, high-resolution, optical flow, and variational Bayes.

**Motoharu Sonogashira** received his B.S. degree in engineering, and M.S. and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan in 2013, 2015, and 2018, respectively. He is currently a researcher at the Academic Center for Computing and Media Studies, Kyoto University. His research interests include image restoration, blind deconvolution, super-resolution, optical flow, and variational Bayes.

**Masaaki Iiyama** received the B.S. degree in engineering informatics, and the M.S. and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2006, respectively. Currently, he is an Associate Professor with the Academic Center for Computing and Media Studies, Kyoto University. His research interests include 3-D shape reconstruction, reflectance and scattering modeling, and computational photography. He is a member of IEEE.