Probabilistic NeRF for 3D Shape Recovery in Scattered Medium

Yoshiki Ono, Fumihiko Sakaue and Jun Sato
Nagoya Insutitute of Technology, Nagoya, Japan
{*y-ono@cv., sakaue@, junsato@*}*nitech.ac.jp*

Keywords: NeRF (Neural Radiance Fields), Ray Tracing, Scattering Medium, Stochastic Gradient Descent.

Abstract: This research proposes a method for analyzing scene information including the characteristics of the medium by representing the space where objects and scattering media such as fog and smoke exist using the NeRF (Neural Radiance Fields) (Mildenhall et al., 2020) representation method of light ray fields. In this study, we focus on the fact that the behavior of rays inside a scattering medium can be expressed probabilistically, and show a method for rendering an image that changes in a probabilistic manner from only a single ray, rather than the entire scattering. By combining this method with a scene representation using the stochastic gradient descent method and a neural network, we show that it is possible to analyze scene information without generating images that directly render light scattering.

1 INTRODUCTION

In recent years, an increasing number of automobiles are equipped with cameras and sensors to acquire information on the surrounding environment. By analyzing the information acquired by these sensors and understanding the surrounding information, safer driving can be achieved. Such methods are usually designed for use in a clear surrounding environment. However, if a scattering medium, such as fog or smoke, is present in the scene, the observed image will be affected by it, resulting in a blurred image(Scadron et al., 1964; Tian et al., 2017). Therefore, it is difficult to obtain appropriate results when processing assumes a clear image. In particular, methods that recover three-dimensional information, such as scene shape reconstruction, have complex ray behavior, making it difficult to recover appropriate information.

In order to eliminate the influence of such scattering medium on cameras and sensors, and to accurately acquire information about the surrounding environment, it is necessary to analyze the optical phenomenon of light scattering that occurs when a ray of light enters the scattering medium. However, inside the scattering medium, light changes its behavior depending on whether or not it impacts on small particles. Therefore, a very complex ray space is formed inside the medium, which is difficult to analyze directly. Various methods have been proposed to solve this problem (Mukaigawa et al., 2010; K.Nayar et al., 2006; Narasimhan et al., 2006; Naik et al., 2015; Kitano et al., 2017; L.G. and J.L., 1941; Satat et al., 2018; Narasimhan et al., 2006)

Nayar (K.Nayar et al., 2006) et al. proposed a method for separating scattered light into a direct component reflected on the object and a global component scattered by the scattering medium using a technique called high-frequency pattern projection. Although this method can be applied to media of various densities, it requires multiple projection of the modulation pattern and multiple imaging of the scene to separate the light rays. Therefore, this method is unsuitable for dynamic scenes.

Narasimhan et al.(Narasimhan et al., 2006) propose a method to estimate the characteristics of the scattering medium itself, but this method requires the condition that only the scattering medium can be measured independently. Satat et al.(Satat et al., 2018) propose a method to analyze the information in the scattering medium obtained by using a sensor to obtain clear information by removing the effects of back scattering. This method makes it possible to obtain information on the surrounding environment from sensors mounted on automobiles and other vehicles in foggy scenes. However, this method is difficult to apply to general scenes because of its limited applicability.

In recent years, a method for analyzing scenes using deep learning has been proposed, but it requires a huge amount of training data to accurately analyze complex scenes. To solve this problem, we propose a

Ono, Y., Sakaue, F. and Sato, J.
Probabilistic NeRF for 3D Shape Recovery in Scattered Medium.
DOI: 10.5220/0012473300003660
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 3: VISAPP*, pages 779-785
ISBN: 978-989-758-679-8; ISSN: 2184-4321
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

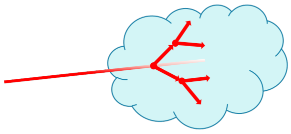


Figure 1: Behavior of rays in scattering medium.

method for describing and estimating scene information, including the state of the scattering medium, using a neural network that does not require prior learning.
As mentioned above, light scattering is a complex phenomenon in which each ray behaves differently in the medium. In this research, instead of representing the entire scattering, we focus on each ray and represent its behavior using a probabilistic model. In this case, although a single ray alone cannot adequately represent a scene, we show that scene information can be adequately estimated by combining it with the stochastic gradient descent method used to train neural networks. This shows that it is possible to analyze scene information with a small amount of computation. We also show that the method can be applied to various applications, such as image generation without the influence of scattering medium.

2 LIGHT SCATTERING

2.1 Scattering Medium

At first, we will explain the scattering medium and the behavior of light inside it. A scattering medium is a medium in which many small particles exist through which light rays travel straight, such as air. Typical scattering media include fog and smoke. As described above, light rays incident on a scattering medium impact the particles in the medium and are reflected in a direction different from their original direction of motion as shown in Fig.1. Furthermore, the reflected ray collides with another particle in the medium, changes direction again, and travels straight ahead. Since the particles inside the medium are very small, the direction of reflection differs greatly for each ray. This results in a mixture of rays traveling in various directions inside the medium. This causes rays of light entering the scattering medium to form a complex ray space inside the medium, which is called light scattering.

2.2 Attenuation

Next, the attenuation of rays by the scattering medium is explained. For this purpose, consider the case where a ray L entering the scattering medium from a certain direction collides with a particle in a straight line in that direction, as shown in Fig. 2. When the ray L collides with a particle in the medium, part of the energy of the incident ray is absorbed by the particle and the intensity of the light decreases. Therefore, the intensity of the ray L' observed at the point $\mathbf{X}$ in the scattering medium decreases with the straight-line distance of the ray. the phenomenon in which the energy of a ray of light is absorbed due to collision with a particle is called light attenuation. This attenuation occurs any time as the light travels straight ahead. Therefore, the amount of attenuation varies with distance. When d is the distance traveled by the light in the medium, the intensity of the attenuated ray L' can be expressed as follows:

$$L' = e^{\sigma_t d} L \tag{1}$$

where σ_t is the attenuation coefficient determined by the density of the medium.

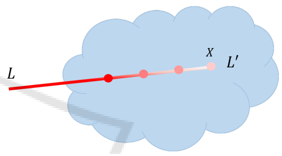


Figure 2: Light attenuation.

2.3 Scattering

Next, we will discuss the scattering of rays of light by the scattering medium, as shown in Fig 3. When a ray of light L entering a medium collides with a particle in the medium, the energy is not only absorbed by the particle, but is also reflected in a different direction from the incident light. This reflection varies depending on the shape and size of the particles. Therefore, a ray of light traveling straight ahead will be transformed into light traveling in a different direction due to minute differences in position and direction. How much light is scattered in the direction θ can be expressed by the phase function as follows:

$$p(\theta) = \frac{1}{4\pi} \cdot \frac{1-g^2}{(1+g^2-2g\cos\theta)^{\frac{2}{3}}} \tag{2}$$

were, $g(-1 \leq g \leq 1)$ is a coefficient that determines the scattering directionality. As shown in Figure 4, when the value of g is 0, the incident light is isotropically scattered in all directions. Furthermore, when the value of g is positive, the scattering direction is forward-scattering, and when the value of g is negative, the scattering direction is backward-scattering.
As described above, light rays incident on a scattering medium travel in various directions within the

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

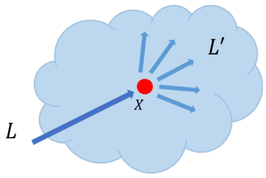


Figure 3: Light scattering.

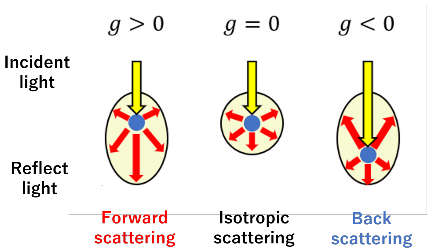


Figure 4: Directionality of the scattering by g.

medium while repeating attenuation and scattering. As a result, light rays traveling in various directions coexist inside the scattering medium, creating a complex ray space.

3 LIGHT FIELD REPRESENTATION BY NeRF

3.1 NeRF

As mentioned in the previous sections, a scene filled with scattering media creates a very complex ray field. Therefore, in order to perform this analysis, a method to express this appropriately is required. In this research, such scene information is expressed using a neural network. Representation of ray fields using neural networks has been used in various ways in recent years, including NeRF. This section provides an overview of how to express scene information using neural networks.

In methods prior to NeRF, representations that sampled multidimensional spaces such as voxels were used to represent 3D scenes and 4D light fields. Although this representation is easily realized, the differential accuracy of scene information depends on the sampling resolution. Therefore, there was a problem that it was insufficient for representing high-dimensional spaces such as ray space. On the other hand, NeRF solves this problem by using a neural network to represent the ray space. In this method, the target scene is represented by a 5-dimensional ray space and this estimation is performed. At this time, when coordinate information in a five-dimensional space is input to the neural network, learning is performed so that the density of the space and the RGB information of the light rays are output. This allows neural networks to be used like continuous lookup tables. By learning so that the output of this neural network matches the input image, scene information can be restored.

3.2 Light Ray Space Estimation and Volume Rendering

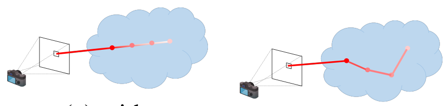
Next, we explain how the ray space is estimated in NeRF. As described in the previous section, in NeRF, the neural network is trained so that the estimated scene information is consistent with the input image, that is, the error between the image generated from the scene information and the input image becomes small. Therefore, it is necessary to render images based on the neural network information.

For this image generation, volume rendering is used. Now, 5-dimensional coordinates are sampled along the camera's ray direction, and each sampling point is converted to RGB values and density using a neural network. The light rays emitted from each point enter the camera with attenuation. In this case, the observed image can be calculated as the sum of these values. By applying this process to all viewpoints and all pixels, an image can be rendered for each viewpoint. This process is easily differentiable because it consists of distance- and density-based attenuation and simple summation. Therefore, the difference between the rendered image and the input image is differentiable as well. By minimizing this error using the gradient-based minimization method, etc., a neural network that appropriately represents scene information can be trained.

4 SCENE INFORMATION RECONSTRUCTION BY SEPARATING SCATTERED LIGHT AND DIRECT LIGHT

4.1 Volume Rendering with Scattering

Using NeRF described in the previous section, scene information can be recovered as a neural network. Since this method targets thick space, scenes with scattering medium such as fog can be restored in the same way. However, the scene information obtained by this method is a mixture of scattered light by the scattering medium and light emitted from objects, and it is difficult to say that the reconstruction results are



(a) with straightforward light behavior

(b) with scattered light behavior

Figure 5: Volume rendering with light scattering.

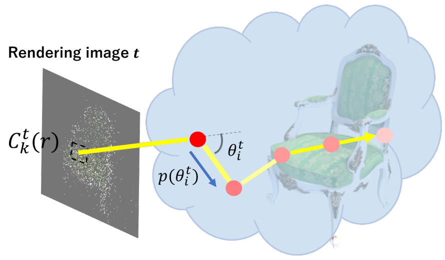


Figure 6: Probabilistic ray tracing.

sufficient for such applications as generating images in which the scattering medium is removed.

For example, consider the case where an object emitting light from its surface exists in a scattering medium and is observed. In order to estimate such a scene where no scattering medium exists, it is necessary to estimate the light directly emitted from the object. However, the result estimated using NeRF is the light field formed after the light irradiated from the object is scattered, which is very different from the light field formed by the light irradiated from the object. Therefore, even if the light rays corresponding to the scattering medium are removed from the recovered result, the scene information obtained from this will be significantly different from the scene information obtained from a scene without a scattering medium. Therefore, if we consider that reconstructing the light field on the surface of an object is equivalent to reconstructing the object, this method cannot properly reconstruct the object inside the scattering medium.

In order to solve such a problem, instead of rendering an image by integrating the light field on a simple straight line, it is necessary to reproduce the scattering of light and estimate only the rays directly irradiated from each point in the scene. In this study, we propose a method of ray tracing that takes light scattering into account when rendering an image, as shown in Fig 5(b). In this case, the rendered image is based on rays of light emitted from each 3D point and arriving at the camera after being scattered. In other words, it is possible to recover the light directly irradiated by each 3D point.

4.2 Probabilistic Ray Tracing

In performing such a rendering, we focus on the property that the path of a ray is the same even if the incoming and outgoing rays are reversed. Using this property, we can achieve scattering-aware rendering by tracking the rays emitted from each pixel while scattering them according to the model described in Section 2. However, even if we track a single ray emitted from a pixel, the ray will be scattered in various directions as shown in Section 2.3. This means that one ray can diverge into multiple rays. Such branches will occur recursively, and it is not practical to keep track of all of them.

In this study, we focus on the probabilistic property of scattering. Considering that the scattering of light is due to the probabilistic behavior of particles when they collide with each other, the spread of light indicated by the phase function is considered to represent the probability of light rays going in each direction. Therefore, as shown in Figure 6, the path of a single ray reaching the camera can be reproduced by tracking the ray at each point in the scattering medium, changing it randomly according to the probability indicated by the phase function.

To achieve such processing, it is necessary to determine whether the point of attention in the scene is a point in the medium, on the surface of the object, or inside the object. Since the density of an object is considered to decrease significantly in a scattering medium, a point whose density estimated in NeRF is less than a threshold is considered to be a scattering medium.

Considering these aspects, the value $C_k^t(t)$ of a certain pixel in the rendered image when only one ray is tracked can be calculated as follows:

$$
\begin{aligned}
C_k^t(t) &= \sum_{i=0}^{N} R_{Xi} && (3)\\
R_{Xi} &= T_i\left(1-\exp\left(-\sigma_i\delta_i\right)\right)\mathbf{c}_{Xi} \\
T_{Xi} &= \exp\left(-\sum_{j=1}^{i-1}\sigma_{Xj}\delta_j\right) \\
X_{i+1} &= X_i+\Delta X_i && (4)
\end{aligned}
$$

where ΔX_i is determined stochastically and distribusion of θ_i^t between ΔX_i and ΔX_{i-1} are represented as follows:

$$
p\left(\theta_i^t\right) = \frac{1}{4\pi}\cdot\frac{1-g^2}{\left(1+g^2-2g\cos\theta_i^t\right)^{\frac{2}{3}}} \tag{5}
$$

In Eq.(4), $\mathbf{c},\sigma$ are the RGB values and density that are output from NeRF, N is the number of sampling points, and δ is the distance between adjacent sampling points. Also, α_s is the scattering coefficient that represents the rate at which light is scattered. As described above, the rendering separates the effects of

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

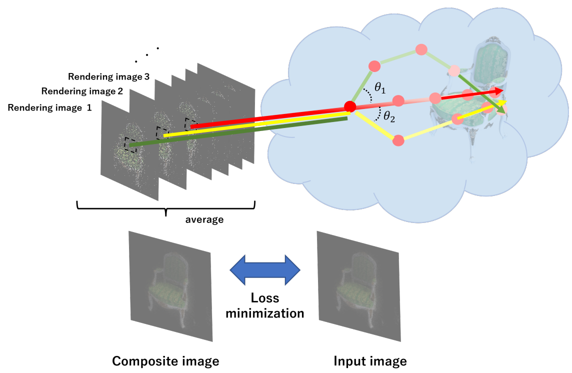


Figure 7: Volume rendering with light scattering.

objects and scattering medium by using different formulas depending on the density threshold d that discriminates between objects and scattering medium.

The image generated as described above does not reproduce scattering because it is not the result of tracking all rays of light. However, as shown in Fig 7, it is possible to generate an image that reproduces light scattering by rendering many similar images and generating an average image. By training the neural network to minimize the difference between this image and the input image, it is possible to analyze scene information with scattering considered.

4.3 Scene Estimation Using Stochastic Gradient Descent

Even with such a method, a large number of images must be generated to render an image that adequately reproduces the scattering. Therefore, a large amount of time is required to train a neural network. Therefore, we focus on the stochastic gradient descent method used in neural network training. Instead of minimizing the loss function calculated from all training images, this method updates the neural network using gradients calculated from randomly selected subsets, called mini-batches. It is known that appropriate learning can be achieved by interchanging these mini-batches.

If we consider the image obtained by tracking only one ray as shown in Fig. 8 as similar to this mini-batch, we can expect to train the neural network appropriately even if we use each of the generated images instead of using the image that reproduces the scattering by averaging. In this case, we can expect to be able to train the neural network appropriately. In this case, the number of generated images can be significantly reduced compared to generating images that reproduce the scattering. Therefore, in this study, the update of the rendering neural network is repeated to reduce the following error function between the

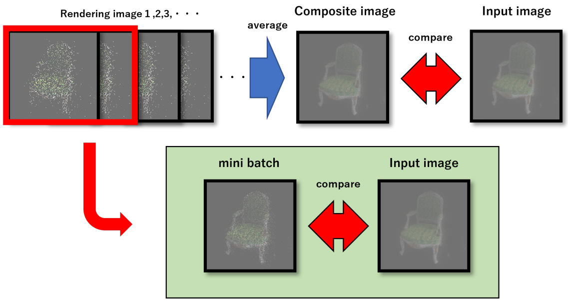


Figure 8: Scene estimation based on Probabilistic image synthesis.

rendered image I_i obtained by stochastic ray tracing and the input image I.

$$Loss = \| I - \frac{1}{t}\sum_{i=0}^{t} I_i \|^2$$

This efficiently recovers only the rays directly irradiated by objects in the scene.

5 EXPERIMENTAL RESULTS

5.1 Environments

We show the results of using the proposed method to restore scene information that eliminates the effects of scattering in a scene where a scattering medium exists. In this experiment, an object was placed in a scattering medium and images taken from various directions were created as simulation data as shown in Fig. 9. To synthesize the training and test image with light scattering, we first synthesized a space with a fog model added with only objects restored using conventional NeRF. Next, by using the proposed rendering method for that space, we created a captured image of the space where the object and scattering medium exist.

In these experiments, 125 images were synthesized, 100 were used as training images for NeRF, and 25 were used as test images. As the target scene, we created a scene in which the upper right chair was illuminated. Figure 10(b) shows an example of the image without a scattering medium. A scattering medium was added to this scene, and images taken in the scattering medium were similarly created. In this image, scattering coefficients $\sigma_s = 0.05$, $g = 0.99$ and $\sigma_s = 0.05$, $g = 0.985$ were used. Figure 11 shows examples of the synthesized images with scattering. We trained a NN using this dataset and evaluated it by comparing images taken from a different viewpoint than the training images and images generated by volume rendering from the trained NN. In addition,

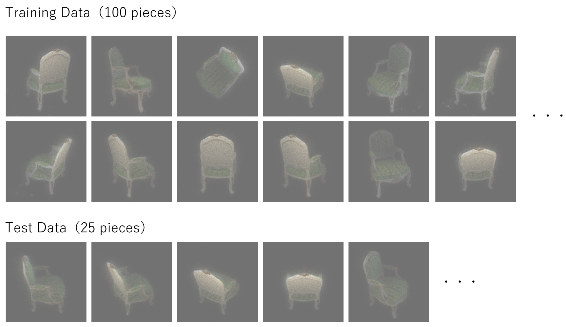
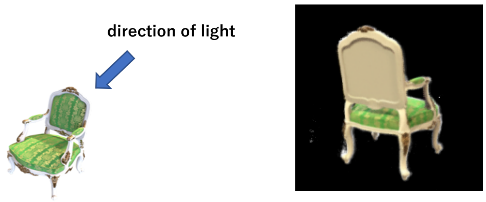


Figure 9: Example images for training and test.



(a) Illuminating direction (b) Synthesized image

Figure 10: Illuminating direction.


(a)without scattering (b)with scattering ($g = 0.985$)

Figure 11: Examples of the input images.

we similarly generated images from the learning data that removed the effects of fog, and compared them as well.

5.2 Results

The results of rendering images containing scattering medium from all scene information recovered using the proposed and conventional methods are shown in Fig. 12(b) and (c). In Fig.12(b), the rendering result was generated as the average of 100 images rendered by probabilistic ray tracing. In (c), the result was rendered from NN including effect of fog directly. For comparison, the ground truth image is shown in Fig. 12(a), and the RMSE with the ground truth image is shown at the bottom of the figure. The results show that both the conventional and the proposed methods are able to reproduce foggy scenes with high accuracy. This confirms that the stochastic ray tracing method described in this paper can appropriately represent light scattering.


(a)Ground truth (b)Proposed NeRF (RMSE:2.168) (c)Normal NeRF (RMSE:2.095)

Figure 12: Estimated results.

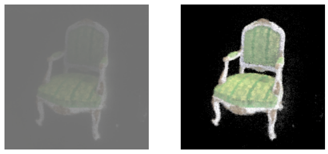

(a) with scattering (b)without scattering

Figure 13: Result of eliminating scattering medium.


(a)Ground truth (b)Proposed NeRF (RMSE:6.205) (c)Normal NeRF (RMSE:9.987)

Figure 14: Comparison between proposed method and normal NeRF.

Next, the result of removing fog from the captured image created by the proposed method is shown in Fig. 13. In this result, the areas of low density estimated by NeRF are judged to be foggy, and the light field in these areas is replaced by 0 in the rendering. For comparison, the same rendering was applied to the results restored using normal NeRF in Fig. 14. The results show that the color of the fog is mixed in with the estimated chair color in the result obtained using normal NeRF, and that the effect of the fog remains in the rendering result. In addition, the fog was not eliminated because of the presence of a certain density of fog around the chairs, and its effect was still rendered. On the other hand, the color of the chair in the rendering result using the proposed method is close to the color of the correct image, confirming that the effect of the fog has been removed. These results confirm that the proposed method can estimate scene information by separating the effects of fog and objects.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

6 CONCLUSION

In this paper, we propose a method for estimating the ray space in which an object and a scattering medium exist simultaneously by utilizing the ray space representation of NeRF and the stochastic characteristics of scattering, and separating the effects of the scattering medium and the object on the ray space. Simulation experiments were conducted on the estimation of ray space and removal of scattering medium using simulation data of scenes in which objects were placed on scattering medium, and the effectiveness of the proposed method was confirmed.

REFERENCES

Kitano, K., Okamoto, T., Tanaka, K., Aoto, T., Kubo, H., Funatomi, T., and Mukaigawa, Y. (2017). Recovering temporal psf using tof camera with delayed light emission. *IPSJ Transactions on Computer Vision and Applications*, 9:9–15.

K.Nayar, S., Krishnan, G., D.Grossberg, M., and Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH 2006 Papers.*, pages 935–944.

L.G., H. and J.L., G. (1941). Diffuse radiation in the galaxy. *The Astrophysical Journal*, 93:70–83.

Mildenhall, B., Srinivasan, P., Tancik, M., Barron, J., Ramamoorthi, R., and Ng, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proc. International Conference on Computer Vision(ECCV 2020).*

Mukaigawa, Y., Yagi, Y., and Raskar, R. (2010). Analysis of light transport in scattering media. In *Proc. Conference on Computer Vision and Pattern Recognitin(CVPR 2010).*

Naik, N., Kadambi, A., Rhemann, C., Izadi, S., Raskar, R., and Kang, S. B. (2015). A light transport model for mitigating multipath interference in time-of-flight sensors. In *Proc. Conference on Computer Vision and Pattern Recognition(CVPR 2015)*, pages 73–81.

Narasimhan, S., Gupta, M., Donner, C., Ramamoorthi, R., Nayar, S., and Jensen, H. (2006). Acquiring scattering properties of participating media by dilution. *ACM Trans. on Graphics*, 25(3):1003–1021.

Satat, G., Tancik, M., and Raskar, R. (2018). Towards photography through realistic fog. In *Proc. International Conference on Computational Photography (ICCP 2018)*, pages 1–10.

Scadron, M., Weinberg, S., and Wright, J. (1964). Functional analysis and scattering theory. *Physical Review*, 135:B202.

Tian, J., Cui, Z. M. T., Zhang, Z., Kriegman, D., and Ramamoorthi, R. (2017). Depth and image restoration from light field in a scattering medium. In *Proc. International Conference on Computer Vision(ICCV 2017)*, pages 2401–2410.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS