Vol.2023-CVIM-233 No.53
2023/3/3

微分可能フォトンマッピングに基づく散乱媒体密度推定

瀬戸口 諒[1]　藤村 友貴[1]　櫛田 貴弘[1]　舩冨 卓哉[1]　向川 康博[1]

概要：本研究では，微分可能フォトンマッピングを用いて多視点画像から散乱媒体の密度を推定する手法を提案する．物理ベースレンダリング手法の 1 つにフォトンマッピングがあり，これは散乱媒体内部で発生する多重散乱をシミュレーションするのに適する．フォトンマッピングを微分可能な形で実装することで，推定画像と正解画像の誤差が小さくなるように散乱媒体の密度を推定した．シミュレーションデータを用いた実験により，推定した密度とレンダリングした画像の定量評価を行った．

1. はじめに

濁った水中や霧，煙などの微粒子が拡散した環境では，入射した光は媒体中の微粒子と衝突し散乱が生じる．このような散乱媒体は実世界ではありふれた存在であり，映画やゲームなどの仮想コンテンツにおいても写実的な表現のためによく用いられる．したがって，実世界において散乱媒体の密度や散乱特性を推定することは，コンピュータビジョンとコンピュータグラフィクスの両分野で長年研究されている．

カメラで撮影された 2 次元画像から散乱媒体の密度を推定する研究が行われている．煙や霧といった流動的な散乱媒体をカメラで撮影したとき，どのような画像が撮影されるかをシミュレーションするには，流体の運動をモデル化することが重要である [3]．このような流体の運動方程式を利用した研究として，Franz ら [4] は，多視点で撮影された散乱媒体の時系列画像から，散乱媒体の密度と時間的な変化を推定する手法を提案している．最適化は微分可能レンダリングを用いて行い，推定された密度からレンダリングした画像と実際の撮影画像との誤差が小さくなるように最適化を行う．

近年は Neural radiance field (NeRF)[9] と同じように，散乱媒体の密度や散乱特性を多層パーセプトロンによって表現し，ボリュームレンダリングによって最適化する手法が提案されている．しかしながら，散乱媒体によって引き起こされる多重散乱 [10] は，カメラの視線方向上だけを考慮したシンプルなボリュームレンダリングで表現することは難しい．2 次の散乱を多層パーセプトロンで近似する研究 [12] や，多重散乱を球面調和関数で近似する研究 [13] も行われているが，近似的にレンダリングを行うこれらの手法では散乱媒体の密度を厳密に推定することは困難であると考えられる．

そこで本研究では，多重散乱を扱うレンダリング手法として，フォトンマッピングの採用を試みる．フォトンマッピングは相互反射や屈折，多重散乱といった複雑な光学現象の再現に長けており，フォトンマッピングを最適化に用いることで，散乱媒体の密度をより厳密に推定することが可能になると考えられる．本研究ではフォトンマッピングを最適化に適用するため，自動微分の機能が提供されている深層学習フレームワークである PyTorch を用いて微分可能な形で実装を行う．多視点で撮影された散乱媒体の画像を入力として，微分可能フォトンマッピングを用いて密度推定を行う手法を提案する．

2. 微分可能レンダリングに基づく散乱媒体密度推定

本章では提案手法の全体像として，散乱媒体の多視点画像を入力とした，微分可能レンダリングによる散乱媒体の密度推定について説明を行う．

2.1 本手法の概要

図 1 に本手法の全体像を示す．入力は散乱媒体を撮影した画像 $\mathbf{y}$ であり，出力は散乱媒体の密度である．対象とする 3 次元空間は 3 次元グリッドによって表現し，各ボクセルにおける密度を推定する．任意の点における密度は線形補完により計算する．Deep reflectance volumes (DRV) [1] と同じように，この 3 次元グリッドは Convolutional

[1] 奈良先端科学技術大学院大学

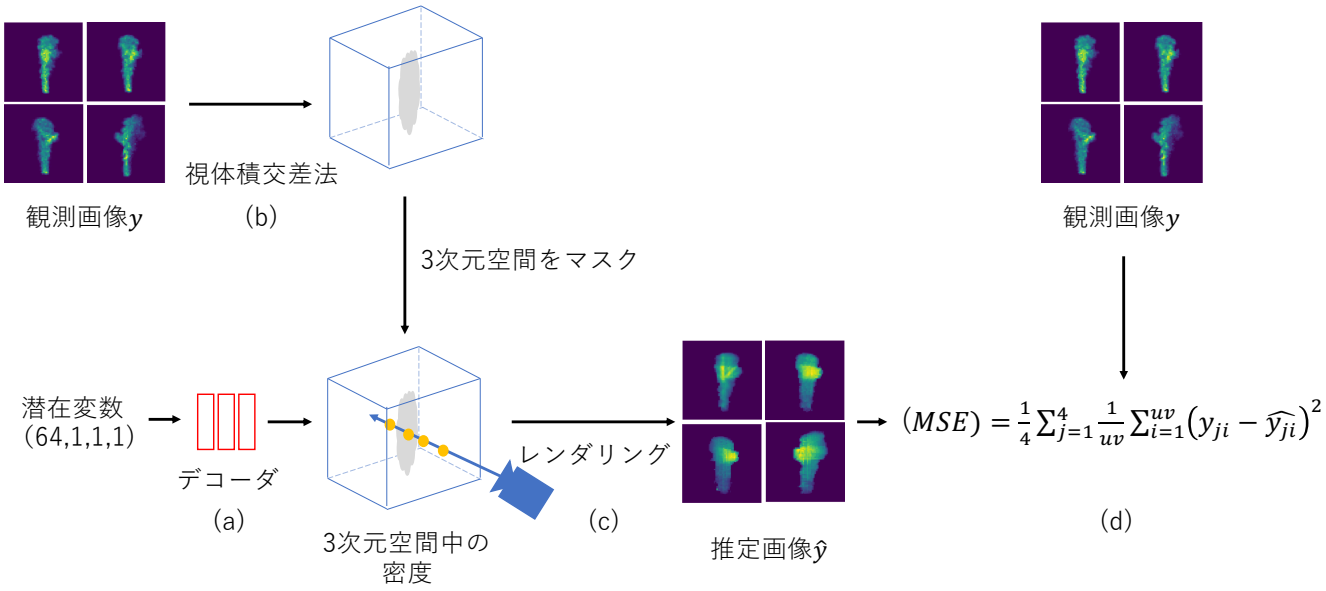

図 1 本手法の全体像．3 次元空間中の散乱媒体の密度は，CNN のデコーダにより 4 次元の潜在変数をアップサンプリングすることで生成される (a)．散乱媒体の密度が 0 の領域は，観測画像における対象のシルエットから求めることができる（b）．3 次元空間中の密度をサンプリングすることで画像を生成する (c)．このレンダラーが微分可能である場合，推定画像と観測画像の誤差を最小化することで潜在変数と CNN のパラメータを最適化することができる（d）．

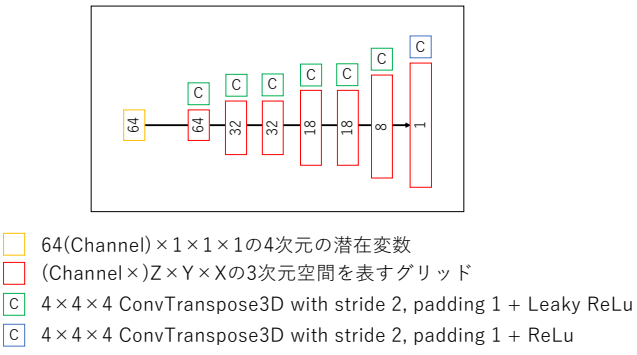

図 2 転置畳み込みによる潜在変数のアップサンプリング

neural network（CNN）で実装されたデコーダの出力として表現し，この CNN への入力である潜在変数と CNN のパラメータを最適化する問題として定式化する．観測画像のみから散乱媒体の密度を推定できるようにするため，CNN が出力した密度から微分可能レンダラーで推定画像 $\hat{\mathbf{y}}$ を生成し，推定画像と観測画像の誤差が小さくなるように最適化を行う．

2.2 散乱媒体密度の生成

散乱媒体の密度は CNN のデコーダに潜在変数を入力することで出力される．**図 2** に実装したネットワーク構造を示す．ネットワークの構造は DRV [1] を参考にした．$64 \times 1 \times 1 \times 1$ の潜在変数を 7 層の 3 次元転置畳み込みでアップサンプリングし，$1 \times 128 \times 128 \times 128$ のグリッドにする．$128 \times 128 \times 128$ のグリッドを $Z \times Y \times X$ の 3 次元空間として扱い，各ボクセルの値がその位置における散乱媒体の密度を表す．

2.3 散乱媒体形状の推定

本研究では空間全体の最適化を行う前に，前処理として視体積交差法により散乱媒体の形状を大まかに推定し，密度が 0 であるボクセルを事前に求める．視体積交差法は，多視点の画像から被写体の 3 次元形状を大まかに推定する手法であり，3 次元空間中の個々の点を各画像平面に逆投影した時，対象のシルエット内部に逆投影される点は残し，それ以外の点は除外する手法である．視体積交差法で得られた散乱媒体の大まかな 3 次元形状を用いて，密度が 0 であるボクセルを事前にマスクすることによって，密度の推定精度が向上すると考えられる．

2.4 微分可能レンダラー

散乱媒体の密度が得られた後は，その密度を用いて画像をレンダリングすることができる．DRV や NeRF では，カメラからの視線方向上のみを考慮したシンプルなボリュームレンダリングによって画像をレンダリングしている．レンダラーを微分可能な形で実装することで，観測画像とレンダリングした画像との誤差を用いて最適化することが可能になる．

散乱媒体内部では，1 度粒子によって散乱された光が別の粒子によって再び散乱される多重散乱が発生する．そのため，DRV や NeRF で用いられているようなシンプルな

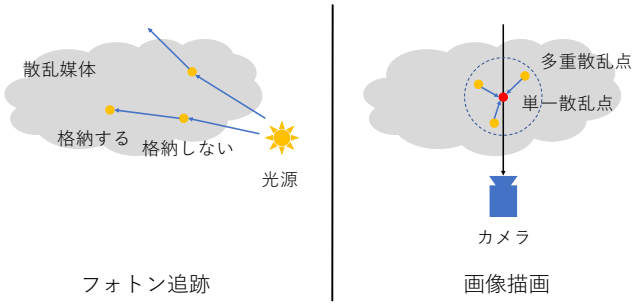

図 3 フォトンマッピングの概観

レンダラーでは霧や煙のような微細粒子状の物質によって引き起こされる光学現象を正確にシミュレーションすることはできない．そこで，本研究ではレンダラーとしてフォトンマッピング [6][7] を用いる．フォトンマッピングは光源から照射されたフォトンを追跡することで，相互反射や屈折といった複雑な光学現象を再現することができるレンダラーであり，散乱媒体下で生じる多重散乱も表現することが可能である．本研究ではこのフォトンマッピングを微分可能な形で実装し，最適化の枠組みに導入する．

3. 微分可能フォトンマッピングにおける近似的な学習

本章では実装したフォトンマッピングの概要と，計算負荷を軽減するための近似的な学習方法について説明を行う．

3.1 フォトンマッピングの実装

本研究では，深層学習フレームワークである PyTorch で提供されている自動微分の機能を用いてフォトンマッピングを実装し，これを微分可能にする．フォトンマッピングの概要を**図 3** に示す．フォトンマッピングは以下のように 2 ステップで画像を生成する．

(1) フォトン追跡　光源からシーン中へフォトンを追跡することによって，フォトンマップを構築する．
(2) 画像描画　フォトンマップ内の情報を使って効率的に画像を描画する．

フォトン追跡

フォトン追跡では，光源からフォトンを照射しシーン中で散乱した位置にフォトンを格納する．このようなフォトンが格納された空間をフォトンマップと呼ぶ．このとき，フォトンが光源から照射され 1 回目に散乱した位置（単一散乱点）は，画像描画時に容易に計算できるため格納しない．散乱媒体内のフォトンは媒体中で影響を受けず通過するか，あるいは媒体と相互作用し散乱か吸収される．散乱されたフォトンの新しい方向は位相関数に基づいた重点的サンプリングを用いて決定する．ここで，位相関数とは光が散乱する方向の分布を表現するものである．最も一般的に利用される位相関数はヘニエイ・グリーンスタインの位相関数 [5] で，これはひずみ係数 $g \in [-1, 1]$ により散乱の指向性を表現する．g の値が 1 に近いほど前方への指向性が強くなり，-1 に近いほど後方への指向性が強くなる．また，$g = 0$ はすべての方向に等しく散乱する等方散乱となる．

画像描画

構築したフォトンマップを使って，散乱媒体の画像を生成する．フォトンマッピングでは，画像に投影される散乱媒体の明るさを単一散乱光と多重散乱光に分けて計算する．単一散乱光の放射輝度はカメラの視線方向上におけるボリュームレンダリングで表せる．光線を長さの小さな区間に分割し，光線漸進法で各区間における単一散乱光をすべて足し合わせる．多重散乱光は光線漸進法の全ての区間について多重散乱により全方位から入射してくる放射輝度を計算することで求めることができる．したがって，シーン中をカメラの視線方向上でサンプリングし，サンプル点における単一散乱光と，その点の周りのフォトンから計算される多重散乱光を計算することで画像を描画することができる．

3.2 近似的な学習法

フォトンマッピングを微分可能な形で実装することで，生成した推定画像と観測画像の誤差を逆伝播することができる．しかし，多重散乱をモデル化した場合は，フォトンはシーン中で複数の散乱を繰り返しカメラに入射するため，全てのフォトンの経路について逆伝播をするには膨大なメモリを要するという問題がある．

光源から散乱媒体を通りカメラに入射するフォトンがどのくらい画像の画素値へ寄与するか，全ての経路について微分を計算し最適化することは計算負荷が大きい．そのため 1 度の学習では，ランダムに選んだいくつかの経路についてだけ微分を計算するようにする．このとき，画像のあるピクセルに到達する経路について，学習に選ばれた経路は学習に選ばれなかった経路の分だけ多く誤差を逆伝播してしまうと考えられる．しかし，学習を繰り返すことで，あるピクセルに到達する全ての経路について等しく学習できると考えられる．

学習する経路の選び方は，単一散乱光についてはカメラからサンプリングした単一散乱点のうち，いくつかの点の経路について微分を計算する．多重散乱光については光源から照射されるフォトンのうち，いくつかのフォトンを追跡しその経路について微分を計算する．こうすることで，1 度の学習でいくつかのフォトンの経路についてだけ推定画像と観測画像の誤差を逆伝播する，近似的な学習が行える．

4. 実験

本研究ではフォトンマッピングで散乱媒体を撮影した画像を作成し，これを正解画像として実験を行なった．有効性の評価として，正解の密度と推定した密度の平均値と

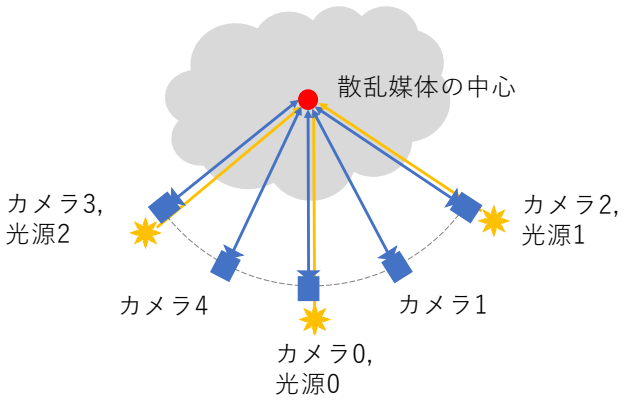

図 4 カメラと光源の配置（上からの視点）

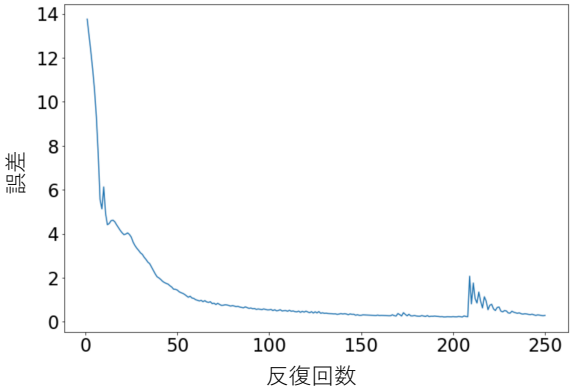

図 5 誤差の推移

最大値を比較し，正解密度と推定密度の散布図を描画した．また，推定した密度で新しい視点から画像をレンダリングし，正解画像との比較を行なった．実験環境は AMD EPYC 7232P@3.1GHz，128GB RAM，NVIDIA GeForce RTX3090 である．

4.1 フォトンマッピングによる画像生成

本研究では散乱媒体の公開データセットである ScalarFlow データセット [2] から密度データを取得し，実験に使う正解画像を作成した．ScalarFlow データセットは多視点で撮影された煙の画像から密度を計測したものである．このとき，0.01 より小さい密度は計測誤差を考慮し 0 に置き換えた．フォトンマッピングにおける消滅係数はこの密度を用いて画像を作成した．散乱媒体は霧を想定し，ヘニエイ・グリーンスタイン位相関数のひずみ係数を $g = 0.9$ [11] とした．多重散乱光は 3 次の散乱まで追跡した．

4.2 カメラと光源の配置

図 4 にシミュレーションのカメラと光源の配置を示す．光源を散乱媒体の中心から同心円上に 3 つ，カメラを散乱媒体の中心から同心円上に 5 台設置された環境を想定してシミュレーションを行った．この時，カメラは散乱媒体の中心を向いているものとした．また，光源は点光源とした．5 台のカメラのうち，4 台を学習用，1 台（カメラ 4）を評価用に使った．

4.3 学習の詳細

損失関数は Mean Squared Error (MSE) を使用し，推定画像と正解画像の画素値の誤差を計算した．単一散乱光はシーン中の単一散乱点のうち，1,000 点を学習に使用した．多重散乱光は光源から照射する 300,000 フォトンのうち，5,000 フォトンを学習に使用した．合計で 250 回のパラメータ更新を行った．パラメータの更新には Adam[8] を用いた．

4.4 実験結果

学習した結果を示す．なお，128 × 128 × 128 の対象とするグリッドに対し，視体積交差法で得られた 68,771 ボクセルでのみ学習を行なった．正解画像と推定画像の画素値の誤差（MSE の値）の推移を**図 5** に示す．誤差は学習を繰り返すほど小さくなっていった．しかし，210 回目あたりの学習で，誤差が大きくなっている．本研究では $g = 0.9$ の前方散乱への強い指向性をもつ散乱媒体を仮定しているため，図 4 のカメラと光源の配置ではカメラに入射する単一散乱光と多重散乱光のほとんどは後方散乱光となる．これらに対し，画像をレンダリングする際に前方散乱光が少しでもカメラに入射すると，それらが画像の輝度値に大きく寄与することでノイズとなってしまう．しかしながら，図 5 に示すように学習途中で一時的に誤差の上昇がみられても，学習をさらに繰り返すことで誤差は収束するという結果が得られた．

次に，学習の反復による推定画像の変化を**図 6** に示す．初期値である学習前の推定画像では，視体積交差法によりシルエットは正解画像と同じだが，画素値のスケールと散乱媒体の濃淡による画素値の大小は正しくない．しかし，反復を繰り返すことで画素値のスケールが合い，正解画像に見られる散乱媒体中央の密度の薄い部分も表現できている，また，正解の画素値と学習後の画素値の差の絶対値であるエラーマップにおいても画素値の差は小さい．これらのことから，学習の反復により潜在変数と CNN のパラメータを最適化できていることがわかる．

次に，学習による密度の平均と最大値の変化を表 1 に示す．視体積交差法で得られた 68，771 ボクセルでの正解密度と推定密度を比較した．学習前，ランダムに生成された密度の平均値と最大値は，正解の密度より小さかった．しかし，学習により潜在変数と CNN のパラメータを最適化することで，出力される密度の平均と最大値が正解に近づいている．このことから，学習により散乱媒体の密度を推定できていることがわかる．

また，正解密度と推定密度の散布図を**図 7** に示す．学習

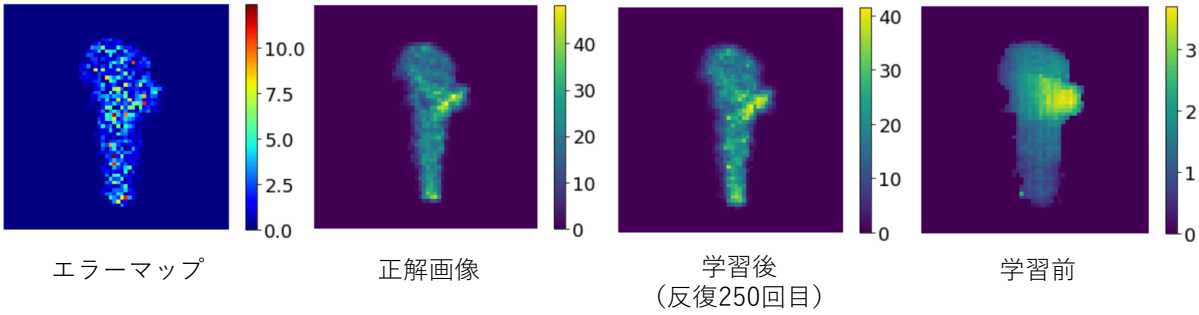

図 6 学習の反復による推定画像の変化（カメラ 2）

表 1 密度の平均値と最大値

	平均値	最大値
学習前の密度	0.0045	0.0209
学習後の密度	0.0417	1.6234
正解の密度	0.0432	1.3585

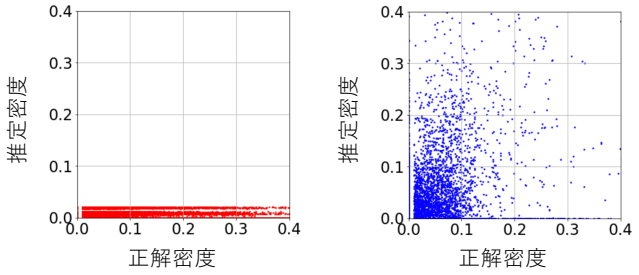

図 7 正解密度と推定密度の散布図

前，ランダムに生成された初期密度は正解密度に対し，無相関である．一方，学習後の推定密度は正解画像に対し，ばらつきは見られるものの正の相関が見られる．このことからも，学習により散乱媒体の密度を推定できていることがわかる．

最後に，推定した密度で新しい視点からレンダリングした画像を正解画像と比較した．**図 8** に示すように，学習に用いていない視点からの画像も正しくレンダリングできていることから，密度が正しく推定できていることがわかる．

5. まとめ

本研究では，微分可能フォトンマッピングにより観測画像のみから散乱媒体の密度を推定する手法を提案した．多重散乱を扱えるフォトンマッピングを微分可能な形で実装した．実験では，散乱媒体の密度の推定と新しい視点からの画像のレンダリングをシミュレーションデータ上で示した．今後は，提案手法の実データへの拡張に取り組んでいきたい．

謝辞 本研究は JSPS 科研費 21K21317 の助成を受けたものである．

参考文献

[1] Bi, S., Xu, Z., Sunkavalli, K., Hašan, M., Hold-Geoffroy, Y., Kriegman, D. and Ramamoorthi, R.: Deep reflectance volumes: Relightable reconstructions from multi-view photometric images, *Proc. European Conf. on Computer Vision (ECCV)* (2020).

[2] Eckert, M.-L., Um, K. and Thuerey, N.: Scalarflow: a large-scale volumetric data set of real-world scalar transport flows for computer animation and machine learning, *ACM Transactions on Graphics (TOG)*, Vol. 38, No. 6, p. 1–16 (2019).

[3] Fedkiw, R., Stam, J. and Jensen, H. W.: Visual simulation of smoke, *In Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH 2001)*, p. 15–22 (2001).

[4] Franz, E., Solenthaler, B. and Thuerey, N.: Global transport for fluid reconstruction with learned selfsupervision, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 1632–1642 (2021).

[5] Henyey, L. G. and Greenstein., J. L.: Diffuse radiation in the galaxy, *Astrophysics Journal* (1941).

[6] Jensen, H. W.: Realistic image synthesis using photon mapping, *AK Peters* (2001).

[7] Jensen, H. W. and Christensen, P. H.: Efficient simulation of light transport in scenes with participating media using photon maps, *In SIGGRAPH 98 Conference Proceedings, Annual Conference Series* (1998).

[8] Kingma, D. P. and Ba, J. L.: Adam: a Method for Stochastic Optimization, *International Conference on Learning Representations (ICLR)*, p. 1–13 (2015).

[9] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R. and Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis, *In Proc. ECCV*, p. 2304–2314 (2020).

[10] Mukaigawa, Y., Yagi, Y. and Raskar, R.: Analysis of light transport in scattering media, *In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 470–476 (2010).

[11] Narasimhan, S. and Nayar., S.: Shedding light on the weather, *In Computer vision and pattern recognition (CVPR)* (2003).

[12] Srinivasan, P. P., Deng, B., Zhang, X., Tancik, M., Mildenhall, B. and Barron, J. T.: NeRV: Neural reflectance and visibility fields for relighting and view synthesis, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 7495–7504 (2021).

[13] Zheng, Q., Singh, G. and Seidel, H.-P.: Neural Relightable Participating Media Rendering, *Advances in*

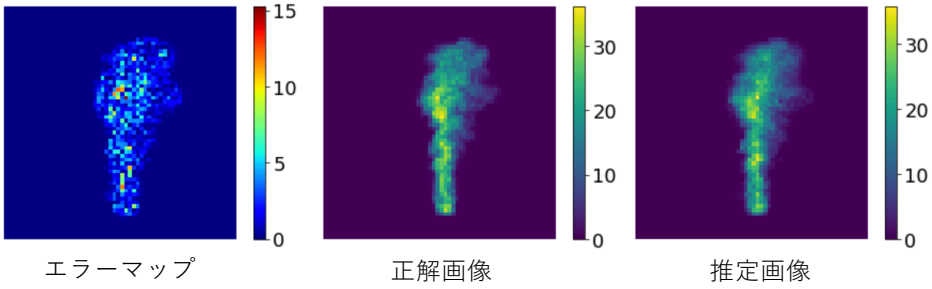

図 8 新しい視点からのレンダリング画像（カメラ 4）

Neural Information Processing Systems 34 (2021).