

Neural Density Field 上での光子追跡による 高時間分解画像の解析手法

藤村 友貴^{1,a)} 櫛田 貴弘^{1,b)} 船富 卓哉^{1,c)} 向川 康博^{1,d)}

概要

本研究では高時間分解画像の新たな解析手法を提案する。近年、neural radiance field (NeRF) に代表される多層パーセプトロン (MLP) を用いたシーンの表現方法が多く提案されている。本研究では neural density field で表現されたシーンにおいて、光子一つ一つを追跡し、single photon avalanche diode (SPAD) カメラで計測された高時間分解画像をレンダリングするように MLP を学習する手法を提案する。これにより、光子一つ一つがどのような経路で SPAD に到達したかを可視化することができる。シミュレーションデータによる実験で、シーンの密度推定と 2 次の相互反射の分離が可能であることを示す。

1. はじめに

カメラで撮影された画像から、光の反射といったシーン中の光の物理現象を解析する研究が行われている。しかしながら、2次元に縮退した通常の画像から相互反射といった複雑な物理現象を解析することは困難である。これに対し、single photon avalanche diode (SPAD) による時間相関単一光子計数 (TCSPC) を利用した研究が行われている [6], [7]。SPAD と高速な時間計測が可能な time-to-digital converter (TDC) を組み合わせたシステムでは、ピコ秒単位の光伝播を計測することができる。近々 SPAD カメラが市場に普及することが予想されており、高時間分解画像の新たな解析手法の検討は非常に重要である。

これに対し、近年深層学習の枠組みによる新たなシーンの表現方法が提案されている。Mildenhall ら [2] は、neural radiance field (NeRF) と呼ばれる多層パーセプトロン (MLP) によるシーンの表現方法を提案している。MLP がシーンの各点における密度と色を推定し、ボリュームレンダリングによって観測画像を再現するように MLP を学習する。

本研究は、高時間分解計測の深層学習を用いた新たな解析手法を提案するものである。具体的には、NeRF のような MLP によるシーン表現を SPAD カメラ による高時間分解計測へ拡張することを試みる。ボリュームレンダリングベースの従来手法 [4] とは異なり、フォトンマッピングに近いアプローチを取る。MLP から出力される密度場 (density field) 上で光子一つ一つを追跡し、SPAD カメラの観測を擬似的に再現する。上記の過程を微分可能な形で実装することで MLP の学習が可能になる。学習した結果を用いることでシーンの密度が得られるとともに、光子一つ一つが相互反射も含めてどのような経路で SPAD に到達したかを可視化することができる。

NeRF ベースの相互反射のモデル化については、1 次の相互反射のみを考慮した従来手法 [5] が存在する。一方で、光子を追跡する本研究の定式化では、相互反射の回数によって計算量が指数関数的に増大することはなく、原理的には任意の n 次の相互反射をモデル化することが可能である。今回実験で用いたデータでは 3 次以上の相互反射の影響は小さかったため、本研究では 2 次の相互反射までをモデル化し、それらの分離と可視化を行なった。

2. 提案手法

2.1 SPAD カメラによる高時間分解計測

SPAD と TDC を用いた計測システムでは、光源から照射された光子一つ一つが確率的に観測される。この観測を複数回繰り返すことで、各時刻 $t \in \{0, \dots, T-1\}$ に到達した光子の数を計測し、ビン数が T のヒストグラム $\mathbf{h} \in \mathbb{R}^T$ を取得できる。SPAD が一般的なカメラのような $H \times W$ の二次元配列である場合は、時間を加えた $H \times W \times T$ の三次元配列が取得できる。

本研究では、確率的な光子の観測を繰り返すというこのプロセスを MLP が出力した密度場上でモデル化し、観測を再現するように MLP を学習する。すべての過程を微分可能な形で実装することで、MLP の学習が可能となる。

2.2 手法の概要

図 1 に本手法の概要を示す。なお、本研究では対象物体

¹ 奈良先端科学技術大学院大学

a) fujimura.yuki@is.naist.jp

b) kushida.takahiro.kh3@is.naist.jp

c) funatomi@is.naist.jp

d) mukaigawa@is.naist.jp

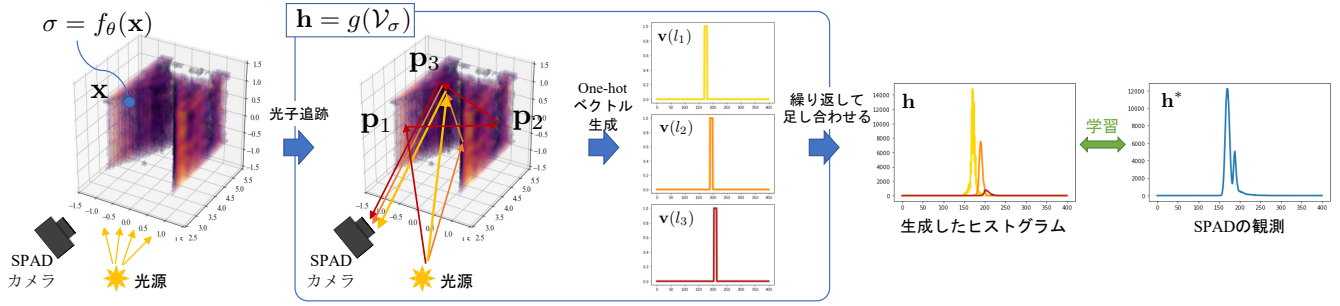


図 1 手法の概要. MLP f_θ に三次元点 \mathbf{x} を入力し密度 σ を求める. 密度場上で光子を追跡して経路長を計算し, 経路長に相当するビン値が 1 である one-hot ベクトルを生成する. この操作を繰り返し足し合わせて SPAD カメラの観測を再現し, f_θ の学習を行う.

は反射率が 1 の完全拡散反射面とし, 2 次の相互反射までをモデル化する.

MLP $f_\theta: \mathbb{R}^3 \rightarrow \mathbb{R}$ は, シーン中の三次元座標 $\mathbf{x} \in \mathbb{R}^3$ を入力として, その点における密度 $\sigma(\mathbf{x}) \in \mathbb{R}$ を出力する. ここで, θ は MLP の学習パラメータである. 本研究では計算量の削減のため, シーンを $81 \times 81 \times 81$ のグリッド状に分割してグリッドの各点を f_θ に入力し, $\mathcal{V}_\sigma \in \mathbb{R}^{81 \times 81 \times 81}$ のボリュームを得る. 任意の点の座標, 密度はバイリニア補間により求める.

出力された \mathcal{V}_σ は $g: \mathbb{R}^{81 \times 81 \times 81} \rightarrow \mathbb{R}^T$ に入力され, ヒストグラム \mathbf{h} を得る. 関数 g は密度場上で光子を追跡することにより SPAD カメラの観測を再現する. f_θ の学習のため, g 内のすべての処理は微分可能な形で実装する必要がある.

点光源から照射された光子は, 密度場上で確率的に計算された点 \mathbf{p}_1 において反射する. その後, \mathbf{p}_1 において次の反射方向をサンプリングし, その反射方向上で次の反射点 \mathbf{p}_2 を計算する. 同様の方法でさらに光子を追跡し, 3 回目の反射点 \mathbf{p}_3 も計算する. そして, 光源 \mathbf{s} から SPAD カメラ \mathbf{c} までの経路長 $l_1 = l(\mathbf{s}, \mathbf{p}_1, \mathbf{c})$, $l_2 = l(\mathbf{s}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{c})$, $l_3 = l(\mathbf{s}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{c})$ を計算する. ここで, $l(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{i=1}^{N-1} \|\mathbf{x}_i - \mathbf{x}_{i+1}\|$ である. これらの経路長から, 各経路長に相当するビン値が 1 である T 次元 one-hot ベクトル $\mathbf{v}(l_1)$, $\mathbf{v}(l_2)$, $\mathbf{v}(l_3)$ を微分可能な形で生成する (式 (4)).

SPAD カメラと同様, 上記のプロセスを複数回繰り返して one-hot ベクトルを足し合わせることで, ヒストグラム \mathbf{h} を生成することができる. 最後に, 生成したヒストグラムと実際に SPAD で観測したヒストグラム \mathbf{h}^* が一致するように f_θ を学習する.

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{(y,x) \sim H \times W} \rho(\mathbf{h}_{y,x}, \mathbf{h}_{y,x}^*) \quad (1)$$

ここで, $\mathbf{h}_{y,x}$, $\mathbf{h}_{y,x}^*$ はそれぞれピクセル (y, x) におけるヒストグラムであり, $\rho(\cdot, \cdot)$ はヒストグラム間の誤差関数である. 本研究ではヒストグラム間のスケールを調整したのち, それらの L1 誤差を計算する. 以降, 各ステップの詳細

細について説明を行う.

2.3 密度場の生成

NeRF [2] と同様, ある三次元点 \mathbf{x} における密度 $\sigma(\mathbf{x})$ は, MLP の出力 $\sigma = f_\theta(\mathbf{x})$ として与えられる. 入力 \mathbf{x} は positional encoding [2] により, 24 次元の空間にあらかじめ埋め込む. その後, 出力の次元数が 64 である全結合層 5 層ののち, 最終層で 1 次元の σ を出力する. 活性化関数には最終層を含めた全ての層で ReLU を用いる.

f_θ から直接出力された $\sigma(\mathbf{x})$ に対して, 物体表面の後ろ側は視認できないことを考慮するため, ある任意の点 \mathbf{x}_0 から見た方向 \mathbf{d} 上の点 $\mathbf{x}_i = \mathbf{x}_0 + i\mathbf{d}$ の密度は以下のように計算する.

$$\sigma_{\mathbf{x}_0, \mathbf{d}}(\mathbf{x}_i) = T_i \sigma(\mathbf{x}_i), \quad T_i = \exp\left(-\sum_{j=0}^{i-1} \sigma(\mathbf{x}_j)\right) \quad (2)$$

T_i は一般的なボリュームレンダリング同様, \mathbf{x}_0 から \mathbf{x}_i の間の透過率を表す.

2.4 密度場からの反射点の推定

ある点 \mathbf{x}_0 から方向 \mathbf{d} へ光子が放射されたとき, 式 (2) で求めた \mathbf{d} 上の密度をサンプリングし, それらから光子の反射点 \mathbf{p} を推定する. いま, サンプリングされた密度が $\{\sigma_{\mathbf{x}_0, \mathbf{d}}(\mathbf{x}_i)\}_{i=1}^N$ であるとする. 本研究では点 \mathbf{x}_i で光子が反射する確率をソフトマックス関数を用いて以下のように計算する.

$$q_i = \frac{\exp(\sigma_{\mathbf{x}_0, \mathbf{d}}(\mathbf{x}_i))}{\sum_j \exp(\sigma_{\mathbf{x}_0, \mathbf{d}}(\mathbf{x}_j))} \quad (3)$$

次の反射点はこれらの確率を用いて $\mathbf{p} = \sum_{i=1}^N q_i \mathbf{x}_i$ として計算する.

しかしながら, 決定的に反射点を計算する上記の方法では, 確率的な SPAD カメラの観測をモデル化することができない. そこで本研究では, 光源から放射された最初の光子の反射点を求める際, ソフトマックスの代わりに確率的に出力を行う Gumbel-softmax [1] を用いる. Gumbel-softmax では, 温度パラメータ τ を小さくするにつれて出力される

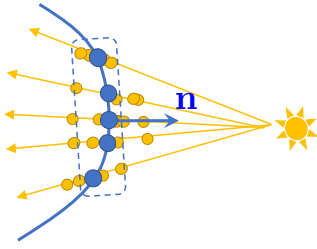


図 2 法線ベクトルの推定. 光源からの各方向上で反射点 (黄点) を平均し, 物体の三次元点群を求める (青点). 各平均点において, 近傍の平均点との外積を計算して法線ベクトルを計算する. 各方向上の反射点の法線ベクトルは, それぞれの方向上の平均点における法線ベクトルを用いる.

値が one-hot ベクトルに近くなる. 本研究では $\tau = 0.1$ として SPAD カメラ の確率的な観測を擬似的に再現する.

2.5 反射方向のサンプリング

本研究では完全拡散反射面を仮定しているため, 次の反射方向は反射点における法線ベクトルを中心とした半球面上からサンプリングを行う. したがって, 次の反射方向を決定するためには, 反射点における法線ベクトルを求める必要がある.

本研究では図 2 に示す方法で法線ベクトルを求める. 3.2 で述べるように, 光源からはある方向に対して複数の光子を照射する. 各光子の反射点は 2.4 で説明したように確率的に決定される (図中黄点). 各方向でこれらの反射点を平均し, 物体の三次元点群を計算する (図中青点). そして, 三次元点群の各点において, その近傍の点を用いて外積により法線ベクトルを求める. 各反射点の法線ベクトルには, それぞれの方向上で求められた法線ベクトルを用いる. 2 回目以降の反射点における法線ベクトルは, 三次元点群の中から距離が最小の点を計算し, その点における法線ベクトルを用いる.

2.6 経路長の one-hot ベクトルへの変換

本研究では 2 次の相互反射をモデル化するため, 光源から最初の反射点 \mathbf{p}_1 を経由して SPAD に到達する経路の長さ l_1 と, 反射点 $\mathbf{p}_1, \mathbf{p}_2$, 及び反射点 $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ を経由して SPAD に到達する経路の長さ l_2, l_3 を計算する. そして, SPAD の観測であるヒストグラムを生成するため, これらの経路長を one-hot ベクトル \mathbf{v} に変換する. 本研究では f_θ の学習のため, このプロセスを以下の式に従って微分可能な形で擬似的に実装する.

$$\mathbf{v}(l) = \frac{\tilde{\mathbf{v}}(l)}{\|\tilde{\mathbf{v}}(l)\|}, \quad \tilde{v}_i(l) = \exp \left(-\frac{(l - l_i)^2}{\gamma} \right) \quad (4)$$

ここで, l_i はヒストグラムのビン i に対応する経路長である. γ はハイパラメータであり, 本研究では $\gamma = 0.5$ とした.

最終的なヒストグラム \mathbf{h} は, 照射した全ての光子の経路

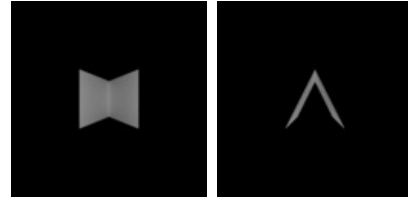


図 3 対象のシーン

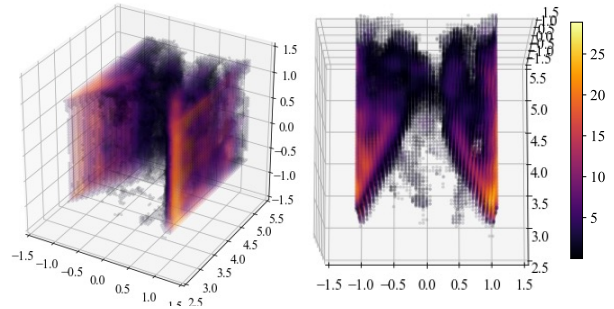


図 4 MLP が推定した密度

についてこれらの one-hot ベクトルを計算し, それら全てを足し合わせて生成する.

3. 実験

3.1 高時間分解画像の取得

本研究ではシミュレーションにより高時間分解画像を生成し, 提案手法の適用を試みた. シミュレーションには MitsubaToFRenderer [3] を用いた. 図 3 は対象物体を正面と上から見た図である. シーン中には反射率が 1 の拡散反射面が二枚あり, その間で相互反射が生じる. 生成した高時間分解画像の画像サイズは 64×64 であり, ヒストグラムのビン数は 400 である. 64×64 ピクセルのうち, あらかじめ物体領域をマスクし, 289 ピクセルのみを学習に用いた.

3.2 光子の照射, ヒストグラム生成

光源はカメラの右手側にあり, 学習時は一回のパラメータ更新につき, 1000 方向をランダムにサンプリングし, 各方向に 1000 個の光子を照射する. したがって, 一度に 1.0×10^6 個の光子を照射する. 各光子がどのピクセルで観測されるかは, 反射点をカメラの内部パラメータで画像上に投影して決定する. 2 回目以降の反射点については, 物体表面で反射したことを保証するため, 2.5 で求めた物体の三次元点群との最小距離が 0.1 以下のものだけをヒストグラムの生成に用いる.

3.3 学習の詳細

学習のバッチサイズは 128 とした. すなわち, 対象とする 289 ピクセルのうちランダムに 128 ピクセルを選択し, これらに対してヒストグラムの誤差を計算してパラメータを一回更新する. 合計で 3000 回パラメータの更新を行

