

# ニューラルインバースレンダリングによる位相シフト法を用いた運動物体の3次元計測の高精度化

浦川 雄気<sup>\*1</sup>      渡辺 義浩<sup>\*1</sup>

**Abstract** - 位相シフト法は構造化光法の一つで、カメラ1台・プロジェクタ1台と最小3枚の正弦波パターンで運動物体の相対位置を計測できる。さらに、複数台のカメラを用いることで、パターン数を増やさずに絶対位置を推定できる。一方、ニューラルインバースレンダリングの導入により、少ない構造化光法のパターンで高精度な計測を実現する手法が新たに提案された。しかし、同手法は物体の動きがレンダリングモデルに組み込まれておらず、運動物体の計測精度が低下する問題があった。また、カメラ1台・プロジェクタ1台を前提としており、位相シフト法に適用した場合、最小パターン数が4枚となる問題があった。そこで、運動物体の高精度な3次元計測に向けた位相シフト法のためのニューラルインバースレンダリングを提案する。具体的には、物体の動きを組み込んだ、カメラ複数台・プロジェクタ1台のレンダリングモデルを構築する。

**Keywords** : 位相シフト法, ニューラルインバースレンダリング, 3次元計測

## 1 はじめに

運動物体の3次元計測は、工場での製品検査、ロボット制御、拡張現実など幅広い分野で応用されている。3次元計測手法の中でも、構造化光法は投影・撮像により計測する手法である [1]。

なかでも、最小3枚の正弦波パターンを投影・撮像する位相シフト法は、高解像度・高精度な計測が可能な手法として知られている [2]。同手法では、絶対位置を求めるアンラッピングと呼ばれる処理が必要であり、そのためにパターンの追加やカメラの複数台化を行う。特に複数台のカメラを用いる場合、パターン数が増えないため、高速に計測できる。しかし、最小でも3枚のパターンを用いるため、パターン間で物体が運動する場合に、位相シフト法の撮像画像モデルと実際の撮像画像の間に誤差が生じる問題がある。

そこで、物体の動きを近似的に撮像画像モデルに導入することで運動物体を計測可能にする3枚パターンの位相シフト法が提案された [3]。しかし、同手法では、撮像画像の各画素について、光輸送係数がパターン間で一定であると仮定する。これにより、撮像画像の各画素に対応する光輸送係数が運動により変化しやすい、複雑な形状などの計測が難しい問題がある。また、少ないパターン数で計測を行い、周囲数ピクセルの画素情報を利用するため、フォトンショット誤差など、撮像画像のランダム誤差の影響を受けやすく、3次元形状の精度が低い問題がある。

一方、少ないパターン数で高精度な計測を行うために、構造化光法にニューラルインバースレンダリング

を導入する手法が新たに提案された [4]。同手法は、構造化光法におけるレンダリングを精密にモデリングし、ニューラルネットワークの高い表現力を活用して3次元形状を大域的に最適化することで、前述のランダム誤差による計測精度の低下を抑える。しかし、物体の動きがレンダリングモデルに組み込まれていないため、運動物体を計測する際に3次元形状に誤差が生じる問題がある。さらに、カメラ1台・プロジェクタ1台の構成に向けたレンダリングモデルであるため、位相シフト法に適用した場合、必要なパターン数が4枚に増える問題がある。

そこで本稿では、運動物体の高精度な3次元計測に向けた位相シフト法のためのニューラルインバースレンダリングを提案する。まず、運動物体の計測のために、物体の位置変化を表す変位場をレンダリングモデルに組み込み、物体の正確な動きを捉える。さらに、運動物体を計測可能な位相シフト法に注目し、3枚のパターンで計測するために、カメラ複数台・プロジェクタ1台でのレンダリングモデルを構築する。

実験では、カメラ2台・プロジェクタ1台の構成で3枚の位相シフト法のパターンを用いた運動物体の計測を達成したことを確認した。さらに、従来の運動物体を計測する位相シフト法 [3] と比較して、計測精度が50%向上したことを確認した。

## 2 関連研究

### 2.1 構造化光法

構造化光法には、複数のパターンを用いるマルチショット法があり、1枚のパターンを用いる手法より計測される3次元形状の解像度が高い。マルチショッ

<sup>\*1</sup>東京科学大学

ト法には、バイナリパターンを投影するグレイコード法、正弦波パターンを投影する位相シフト法 [2]、静止物体の計測に最適化されたパターンを投影する A La Carte 法 [5] がある。しかし、複数のパターンを用いるため、パターン間で物体が運動する場合に、撮像画像モデルと実際の撮像画像の間に誤差が生じる問題がある。

そこで、運動物体の計測のために、動きを近似的に撮像画像モデルに導入する 3 枚パターンの位相シフト法が提案された [3]。しかし、少ないパターン数で計測を行い、周囲数ピクセルの画素情報を利用するため、フォトンショット誤差など、撮像画像のランダム誤差の影響を受けやすい問題がある。

## 2.2 ニューラルインバースレンダリング

構造化光法において、少ないパターン数で高精度な計測を実現するために、ニューラルインバースレンダリングを導入する手法が提案された [4]。同手法は、構造化光法におけるレンダリングを精密にモデリングすることで実現され、構造化光法の特徴的なパターン投影と、ニューラルネットワークの高い表現力による 3 次元形状の大域的最適化を組み合わせた手法である。これにより、カメラ 1 台・プロジェクタ 1 台の構成において、3～4 枚の少ない数の構造化光法のパターンを用いて高精度な 3 次元形状を計測できる。

しかし、パターン間で物体が運動する場合に、動きがレンダリングモデルに組み込まれていないため、3 次元形状に誤差が生じる問題がある。また、位相シフト法に適用する場合、カメラ 1 台・プロジェクタ 1 台の構成では、必要なパターン数が 4 枚に増加する問題がある。

## 3 予備知識

### 3.1 位相シフト法

位相シフト法では、プロジェクタとカメラの画素対応を位相を用いて求め、3 次元計測を行う。まず、次式のような投影画素値  $I_n^p(\mathbf{p})$  の正弦波パターンを投影する。

$$I_n^p(\mathbf{p}) = I_s^p[1 + \cos(\phi(\mathbf{p}) + \theta_n)] \quad (1)$$

ここで、 $n$  は番号、 $\mathbf{p} = (x^p, y^p)$  は投影画像の画素座標、 $I_s^p$  は投影光の強度、 $\phi$  は位相、 $\theta_n$  は位相シフト量である。運動物体を計測するために、Weise らはこの投影画像に対し、次式のような近似された撮像画像モデルを提案した [3]。

$$I_n^c(\mathbf{c}) = I^e(\mathbf{c}) + \alpha(\mathbf{c})I_s^p[1 + \cos(\phi'(\mathbf{c}) + \theta_n + n\Delta)] \quad (2)$$

$$\phi(\mathbf{c}) = \phi'(\mathbf{c}) + 2\pi k(\mathbf{c}) \quad (3)$$

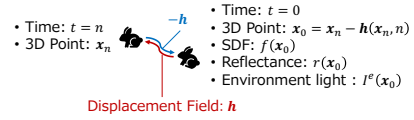


図 1 変位場の概要。

Fig. 1 Overview of the displacement field.

ここで、 $I^e$  は環境光、 $\mathbf{c} = (x^c, y^c)$  は撮像画像の画素座標、 $\alpha$  は光輸送係数、 $\phi'$  はラップされた位相、 $\Delta$  は運動を表すパラメータ、 $k$  は周期数である。同モデルでは、撮像画像の各画素について、光輸送係数  $\alpha(\mathbf{c})$  がパターン間で一定であると仮定する。これにより、撮像画像の各画素に対応する光輸送係数が運動により変化しやすい、複雑な形状などの計測が難しい問題がある。さらに、位相  $\phi'$  は  $0 \leq \phi' < 2\pi$  にラップされているため、絶対位置を求めるために周期数  $k$  を求めるアンラッピングが必要となる。このためにパターンの追加やカメラの複数台化を行う。

### 3.2 変位場

Pumarola らは、変位場を用いることで、複数視点・複数時刻の撮像画像から放射場を推定する手法を提案した [6]。時刻  $t$  における 3 次元点  $\mathbf{x}_t$  の物体は、次式のような時刻  $t=0$  における 3 次元点  $\mathbf{x}_0$  の物体と対応していると考えられる。

$$\mathbf{x}_0 = \mathbf{x}_t - \Delta\mathbf{x}_t \quad (4)$$

ここで、 $\Delta\mathbf{x}_t$  は時刻  $t$  における物体の位置変化を表す。さらに、 $\Delta\mathbf{x}_t = \mathbf{h}(\mathbf{x}_t, t)$  となる関数  $\mathbf{h}$  を導入することで、式 (4) を次式で表現できる。

$$\mathbf{x}_0 = \mathbf{x}_t - \mathbf{h}(\mathbf{x}_t, t) \quad (5)$$

すなわち、関数  $\mathbf{h}$  は物体の動きを表す変位場である。この表現は、複数視点・複数時刻の撮像画像からの放射場や 3 次元形状の推定に有効である。

### 3.3 ニューラルインバースレンダリングによる構造化光法

Mirdehghan らは、カメラ 1 台・プロジェクタ 1 台の構成におけるレンダリングをモデル化した [4]。同手法では、A La Carte 法や位相シフト法などの構造化光法のパターンを使用して、3～4 枚で 3 次元形状を計測できる。これに対し、提案手法では物体の動きを組み込んだ、カメラ複数台・プロジェクタ 1 台のレンダリングモデルを構築する。Mirdehghan らのレンダリングモデルによる撮像画像と投影画像の関係を以下に示す。

まず、投影画像・撮像画像の画素から射出される光線上の 3 次元点は次式で表される。

$$\mathbf{x}(\mathbf{u}, \delta) = \mathbf{o} + \delta d(\mathbf{u}) \quad (6)$$

ここで、 $\mathbf{o}$  はカメラ・プロジェクタの光学中心、 $\mathbf{d}$  は光線の方向ベクトル、 $\mathbf{u} \in \{\mathbf{c}, \mathbf{p}\}$  は撮像画像・投影画像の画素座標である。さらに、投影画像  $I_n^p$  から撮像画像  $\hat{I}_n^c$  を次式のように表現できる。

$$\begin{aligned} \hat{I}_n^c(\mathbf{c}) &= \int_{\delta_{\min}}^{\delta_{\max}} [I^e(\mathbf{x}(\mathbf{c}, \delta)) + r(\mathbf{x}(\mathbf{c}, \delta))] \\ &\quad \cdot \langle \boldsymbol{\eta}(\mathbf{c}, \delta), \mathbf{i}(\mathbf{x}(\mathbf{c}, \delta)) \rangle \\ &\quad \cdot (B * I_n^p)(\mathbf{p}(\mathbf{x}(\mathbf{c}, \delta))) tr(\mathbf{c}, \delta) \sigma(\mathbf{x}(\mathbf{c}, \delta)) d\delta \end{aligned} \quad (7)$$

$$\boldsymbol{\eta}(\mathbf{c}, \delta) = \frac{\nabla f(\mathbf{x}(\mathbf{c}, \delta))}{\|\nabla f(\mathbf{x}(\mathbf{c}, \delta))\|} \quad (8)$$

$$\sigma(\mathbf{x}(\mathbf{c}, \delta)) = \max\left(\frac{-\frac{d}{d\delta}\Phi_s(f(\mathbf{x}(\mathbf{c}, \delta)))}{\Phi_s(f(\mathbf{x}(\mathbf{c}, \delta)))}, 0\right) \quad (9)$$

$$\Phi_s(x) = \frac{1}{1 + \exp(-sx)} \quad (10)$$

$$tr(\mathbf{c}, \delta) = \exp\left(-\int_{\delta_{\min}}^{\delta} \sigma(\mathbf{x}(\mathbf{c}, \delta')) d\delta'\right) \quad (11)$$

ここで  $[\delta_{\min}, \delta_{\max}]$  はレンダリング範囲、 $r$  は反射率、 $\boldsymbol{\eta}$  は表面法線、 $f$  は SDF、 $\mathbf{i}$  は 3次元点  $\mathbf{x}$  からプロジェクタへの単位ベクトル、 $B$  はぼけカーネル、 $\langle \cdot \rangle$  は内積演算子、 $*$  は畳み込み演算子、 $\sigma$  は密度、 $\Phi_s$  はパラメータ  $s$  を用いて、SDF から密度への変換に利用する関数 [7]、 $tr$  は透過率である。さらに、 $\mathbf{p}(\mathbf{x})$  は 3次元点  $\mathbf{x}$  をプロジェクタの投影面に射影した際の投影画像の画素座標を表す。また、拡散反射を想定しており、表面法線と 3次元点からプロジェクタへの単位ベクトルの内積と、反射率  $r$  の積が式 (2) の光輸送係数  $\alpha$  に相当する。

一方、撮像画像  $I_n^c$  から投影画像  $\hat{I}_n^p$  を次式のように表現できる。

$$\begin{aligned} B * \hat{I}_n^p(\mathbf{p}) &= \int_{\delta_{\min}}^{\delta_{\max}} \frac{I_n^c(\mathbf{c}(\mathbf{p}, \delta)) - I^e(\mathbf{x}(\mathbf{p}, \delta))}{r(\mathbf{x}(\mathbf{p}, \delta)) \langle \boldsymbol{\eta}(\mathbf{p}, \delta), \mathbf{i}(\mathbf{x}(\mathbf{p}, \delta)) \rangle} \\ &\quad \cdot tr(\mathbf{p}, \delta) \sigma(\mathbf{x}(\mathbf{p}, \delta)) d\delta \end{aligned} \quad (12)$$

ここで、 $\mathbf{c}(\mathbf{x})$  は 3次元点  $\mathbf{x}$  をカメラの撮像面に射影した際の撮像画像の画素座標を表す。

さらに反射率  $r$ 、環境光  $I^e$ 、SDF  $f$  をニューラルネットワークで表現し、構造化光法に対するニューラルインバースレンダリングを実現する。また、反射率  $r$ 、環境光  $I^e$  は、引数の 3次元点  $\mathbf{x}$  をカメラの撮像面に射影し、ニューラルネットワークへの入力を 2次元にする。このレンダリングモデルを用いることで、高精度な 3次元計測が可能となる。しかし、カメラ 1台・プロジェクタ 1台の構成であるため、位相シフト法のパターンでは最小 4 枚となる。我々の再現実験では、3枚の場合、学習が発散したことを確認した。

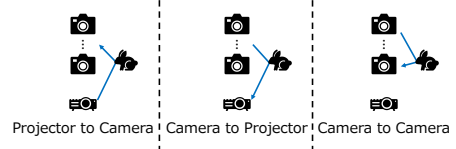


図 2 レンダリングモデルの概要。  
Fig. 2 Overview of the rendering model.

#### 4 運動物体のための位相シフト法へのニューラルインバースレンダリングの導入

本手法では Mirdehghan らの手法 [4] に着想を得て、構造化光法におけるレンダリングモデルを拡張する。まず、運動物体の計測のために、物体の位置変化を表す変位場を組み込む。さらに、運動物体を計測可能な位相シフト法に注目し、パターン 3 枚で計測するために、カメラを複数台化したレンダリングモデルを構築する。変位場の概要を図 1 に、レンダリングモデルの概要を図 2 に示す。

##### 4.1 運動物体に対するカメラ複数台・プロジェクタ 1 台のレンダリング

まず、パターンの番号  $n$  を時刻とみなし、式 (5) を用いることで、変位場を表す。さらに、カメラ複数台・プロジェクタ 1 台の構成において、投影画像・撮像画像に対して 3 種類の推定を行う。

1 つ目はプロジェクタの投影画像からカメラ  $i$  の撮像画像の推定である。式 (7) を拡張することで、プロジェクタの投影画像の画素値  $I_n^p$  からカメラ  $i$  の撮像画像の画素値  $\hat{I}_n^{c_i}$  を次式のように表現できる。

$$\begin{aligned} \hat{I}_n^{c_i}(\mathbf{c}_i) &= \int_{\delta_{\min}}^{\delta_{\max}} [I^e(\mathbf{x}_0(\mathbf{c}_i, \delta, n)) + r(\mathbf{x}_0(\mathbf{c}_i, \delta, n))] \\ &\quad \cdot \langle \boldsymbol{\eta}_n(\mathbf{c}_i, \delta, n), \mathbf{i}(\mathbf{x}_n(\mathbf{c}_i, \delta)) \rangle \\ &\quad \cdot (B * I_n^p)(\mathbf{p}(\mathbf{x}_n(\mathbf{c}_i, \delta))) tr'(\mathbf{c}_i, \delta, n) \sigma(\mathbf{x}_0(\mathbf{c}_i, \delta, n)) d\delta \end{aligned} \quad (13)$$

$$\mathbf{x}_n(\mathbf{c}_i, \delta) = \mathbf{o} + \delta \mathbf{d}(\mathbf{c}_i) \quad (14)$$

$$\mathbf{x}_0(\mathbf{c}_i, \delta, n) = \mathbf{x}_n(\mathbf{c}_i, \delta) - \mathbf{h}(\mathbf{x}_n(\mathbf{c}_i, \delta), n) \quad (15)$$

$$\boldsymbol{\eta}_n(\mathbf{c}_i, \delta, n) = R_n(\mathbf{c}_i, \delta, n) \boldsymbol{\eta}_0(\mathbf{c}_i, \delta, n) \quad (16)$$

$$\boldsymbol{\eta}_0(\mathbf{c}_i, \delta, n) = \frac{\nabla f(\mathbf{x}_0(\mathbf{c}_i, \delta, n))}{\|\nabla f(\mathbf{x}_0(\mathbf{c}_i, \delta, n))\|} \quad (17)$$

$$R_n(\mathbf{c}_i, \delta, n) = \text{Rotate}\left(\frac{\partial}{\partial \delta} \mathbf{x}_0(\mathbf{c}_i, \delta, n), \frac{\partial}{\partial \delta} \mathbf{x}_n(\mathbf{c}_i, \delta)\right) \quad (18)$$

$$tr'(\mathbf{c}_i, \delta, n) = \exp\left(-\int_{\delta_{\min}}^{\delta} \sigma(\mathbf{x}_0(\mathbf{c}_i, \delta', n)) d\delta'\right) \quad (19)$$

ここで、 $\mathbf{x}_n$  は時刻  $t = n$  における 3次元点、 $\mathbf{x}_0$  は時刻  $t = 0$  における 3次元点、 $\boldsymbol{\eta}_n$  は時刻  $t = n$  における

3次元点  $\mathbf{x}_n$  の法線,  $\boldsymbol{\eta}_0$  は時刻  $t=0$  における3次元点  $\mathbf{x}_0$  の法線,  $tr'$  は番号  $n$  における透過率である。また,  $R_n$  は時刻  $t=0$  における3次元点  $\mathbf{x}_0$  での法線から, 時刻  $t=n$  における3次元点  $\mathbf{x}_n$  での法線への回転行列である。この回転行列  $R_n$  を求めるために, まず各時刻の3次元点に対するパラメータ  $\delta$  の勾配ベクトルを求める。さらに, それらの勾配ベクトルが対応していると考え, 回転行列を求める関数 Rotate を用いて計算する。加えて, 反射率  $r$ , 環境光  $I^e$ , SDF  $f$  に関する関数は時刻  $t=0$  での3次元点  $\mathbf{x}_0$  が引数となる。

2つ目はカメラ  $i$  の撮像画像からプロジェクタの投影画像の推定である。式 (12) を拡張することで, カメラ  $i$  の撮像画像の画素値  $I_n^{c_i}$  から投影画像の画素値  $\hat{I}_n^p$  を次式のように表現できる。

$$\begin{aligned} & B * \hat{I}_n^p(\mathbf{p}) \\ &= \int_{\delta_{\min}}^{\delta_{\max}} \frac{I_n^{c_i}(\mathbf{c}_i(\mathbf{x}_n(\mathbf{p}, \delta, n))) - I^e(\mathbf{x}_0(\mathbf{p}, \delta, n))}{r(\mathbf{x}_0(\mathbf{p}, \delta, n)) \langle \boldsymbol{\eta}_n(\mathbf{p}, \delta, n), \mathbf{i}(\mathbf{x}_n(\mathbf{p}, \delta, n)) \rangle} \\ & \quad \cdot tr'(\mathbf{p}, \delta, n) \sigma(\mathbf{x}_0(\mathbf{p}, \delta, n)) d\delta \quad (20) \end{aligned}$$

ここで,  $\mathbf{c}_i(\mathbf{x})$  は3次元点  $\mathbf{x}$  をカメラ  $i$  の撮像面に射影した際の撮像画像の画素座標を表す。

3つ目はカメラ  $j$  の撮像画像からカメラ  $i$  の撮像画像の推定である ( $i \neq j$ )。これは従来手法から新しく追加されたものである。カメラ  $j$  の撮像画像の画素値  $I_n^{c_j}$  からカメラ  $i$  の撮像画像の画素値  $\tilde{I}_n^{c_i}$  を次式のように表現できる。

$$\begin{aligned} & \tilde{I}_n^{c_i}(\mathbf{c}_i) \\ &= \int_{\delta_{\min}}^{\delta_{\max}} I_n^{c_j}(\mathbf{c}_j(\mathbf{x}_n(\mathbf{c}_i, \delta))) tr'(\mathbf{c}_i, \delta, n) \sigma(\mathbf{x}_0(\mathbf{c}_i, \delta)) d\delta \quad (21) \end{aligned}$$

反射率  $r$ , 環境光  $I^e$ , SDF  $f$ , 変位場  $\mathbf{h}$  をニューラルネットワークで表現し, 運動物体に向けた位相シフト法へのニューラルインバースレンダリングの導入を実現する。また, 反射率  $r$ , 環境光  $I^e$  は時刻  $t=0$  の3次元点  $\mathbf{x}_0$  をプロジェクタの投影面に射影し, ニューラルネットワークへの入力を2次元にする。投影面に射影することで, プロジェクタで投影される位置の3次元点と射影された画素座標は1対1で対応する。

## 4.2 最適化

学習のための損失関数について述べる。まず, 4.1節で求めた推定画素値と実際の画素値の差について, 次式のようなプロジェクタ to カメラ損失関数  $\mathcal{L}_{\text{pro2cam}}$ , カメラ to プロジェクタ損失関数  $\mathcal{L}_{\text{cam2pro}}$ , カメラ to カメラ損失関数  $\mathcal{L}_{\text{cam2cam}}$  を用いる。

$$\mathcal{L}_{\text{pro2cam}} = \sum_{i, \mathbf{c}_i, n} \|I_n^{c_i}(\mathbf{c}_i) - \hat{I}_n^{c_i}(\mathbf{c}_i)\| \quad (22)$$

$$\begin{aligned} & \mathcal{L}_{\text{cam2pro}} \\ &= \sum_{\mathbf{p}, n} \|w_{n, \text{cam2pro}}(\mathbf{p}) [B * I_n^p(\mathbf{p}) - B * \hat{I}_n^p(\mathbf{p})]\| \quad (23) \end{aligned}$$

$$\begin{aligned} & w_{n, \text{cam2pro}}(\mathbf{p}) \\ &= \int_{\delta_{\min}}^{\delta_{\max}} r(\mathbf{x}_0(\mathbf{p}, \delta, n)) \langle \boldsymbol{\eta}_n(\mathbf{p}, \delta, n), \mathbf{i}(\mathbf{x}_n(\mathbf{p}, \delta, n)) \rangle \\ & \quad \cdot tr'(\mathbf{p}, \delta, n) \sigma(\mathbf{x}_0(\mathbf{p}, \delta, n)) d\delta \quad (24) \end{aligned}$$

$$\mathcal{L}_{\text{cam2cam}} = \sum_{i, \mathbf{c}_i, n} \|I_n^{c_i}(\mathbf{c}_i) - \tilde{I}_n^{c_i}(\mathbf{c}_i)\| \quad (25)$$

ここで,  $w_{\text{cam2pro}}$  は式 (20) について, 反射率  $r$  や, 表面法線  $\boldsymbol{\eta}_n$  と時刻  $t=n$  の3次元点  $\mathbf{x}_n$  からプロジェクタへの単位ベクトル  $\mathbf{i}$  の内積が小さい場合に, 撮像画像に含まれるランダム誤差が拡大されることを防ぐ。また, 実際の学習では, 距離関数  $\|\cdot\|$  について, L1 ノルムと L2 ノルムの両方を使用する。

さらに, 撮像画像に写らない3次元形状を除去するために, 次式のようなマスク損失関数  $\mathcal{L}_{\text{mask}}$  を用いる。

$$\mathcal{L}_{\text{mask}} = \sum_{i, \mathbf{c}_i, n} \text{BCE} \left( \mathcal{M}^{c_i}(\mathbf{c}_i), O^{c_i}(\mathbf{c}_i) \right) \quad (26)$$

$$\mathcal{M}^{c_i}(\mathbf{c}_i) = \begin{cases} 1 & \text{if } Th^{c_i}(\mathbf{c}_i) \geq \xi \\ 0 & \text{if } Th^{c_i}(\mathbf{c}_i) < \xi \end{cases} \quad (27)$$

$$Th^{c_i}(\mathbf{c}_i) = \max_n I_n^c(\mathbf{c}_i) - \min_n I_n^c(\mathbf{c}_i) \quad (28)$$

$$O^{c_i}(\mathbf{c}_i) = \int_{\delta_{\min}}^{\delta_{\max}} tr'(\mathbf{c}_i, \delta, n) \sigma(\mathbf{x}_0(\mathbf{c}_i, \delta, n)) d\delta \quad (29)$$

ここで, BCE はバイナリクロスエントロピーである。また,  $\mathcal{M}_{c_i}$  はカメラ  $i$  の撮像画像のバイナリマスクであり, カメラ  $i$  の全ての撮像画像の各画素の画素値の最大値と最小値の差  $Th^{c_i}$  に閾値  $\xi$  をかけて作成する。

最後に, SDF の勾配の大きさが1になるようにエイコナル損失関数  $\mathcal{L}_{\text{eik}}$  を, また SDF が0となる3次元点を正則化するためにスパース損失関数  $\mathcal{L}_{\text{sp}}$  を用いる。これらは次式で表される。

$$\mathcal{L}_{\text{eik}} = \sum_{i, \mathbf{c}_i, \delta, n} (\|\nabla f(\mathbf{x}_0(\mathbf{c}_i, \delta, n))\|_2 - 1)^2 \quad (30)$$

$$\mathcal{L}_{\text{sp}} = \sum_{i, \mathbf{c}_i, \delta, n} \exp(-|f(\mathbf{x}_0(\mathbf{c}_i, \delta, n))|) \quad (31)$$

## 5 実装詳細

Multiresolution Hash Encoding [8] を利用したニューラルインバースレンダリングの公開ソースコード [9] と, Mirdehghan らの構造化光法にニューラルインバースレンダリングを導入する手法 [4] に基づいて実装した。変位場への逆伝播を安定させるためには, Multiresolution Hash Encoding の補間を滑らかにする必要があるので [10], Smooth step 補間を採用した。



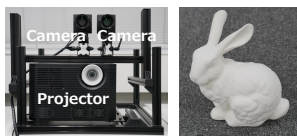


図3 システム構成 (左) と計測対象 (右).  
Fig.3 System (left) and target (right).

また、位相シフト法の正弦波パターンを3枚、カメラを2台用いた。使用したGPUはRTX 3090で、反復回数は合計10,000回行った。学習全体には約23分要した。安定した学習のため、以下の操作を行った。まず、最初の1,000回は変位場  $\mathbf{h} = \mathbf{0}$  とし、残りの9,000回で変位場  $\mathbf{h}$  の学習を進めた。これはパターン枚数が少ないため、変位場  $\mathbf{h} = \mathbf{0}$  とした場合でも、学習が可能であることを利用した。また、式(22)のプロジェクト to カメラ損失関数  $\mathcal{L}_{\text{pro2cam}}$  は、絶対位置を大まかに推定した後に学習させるため、最初の200回は導入しなかった。さらに、式(23)のカメラ to プロジェクト損失関数  $\mathcal{L}_{\text{cam2pro}}$  は、式(20)の割り算による不安定性を回避するため、最初の1,000回は導入しなかった。加えて、式(27)で作成したマスク  $\mathcal{M}^{ci}$  について、物体の境界付近に対応する撮像画像の画素について考える。この画素では、マスク  $\mathcal{M}^{ci}$  が1であっても、ある時刻では画素上に物体がない可能性があり、マスクは正確ではない。そこで式(26)のマスク損失関数  $\mathcal{L}_{\text{mask}}$  は最初の1,500回のみ導入した。

さらに、学習の重みについては、式(22)のL1ノルム、L2ノルムはそれぞれ0.09, 0.91, 式(23)のL1ノルム、L2ノルムはそれぞれ0.09, 0.91, 式(25)のL1ノルム、L2ノルムはそれぞれ0.09, 0.91, 式(26), (30), (31)はそれぞれ0.1, 0.1, 0.01とした。ぼけカーネル  $B$  のサイズは  $1 \times 11$  pixelとした。さらにSDFを用いて、すべてのカメラから見える部分のメッシュを作成し、これを3次元形状とした。

## 6 実験

### 6.1 実験条件

実環境では、デバイスとして、解像度  $1024 \times 768$  pixel のカメラ EoSens 1.1CXP2 (MC1166, monochrome) を2台、解像度  $1024 \times 768$  pixel のプロジェクタ [11] を1台使用した。カメラとプロジェクタは同期されており、フレームレートは500 fpsであった。使用したシステムを図3の左に示す。また、式(1)の正弦波パターンは、周期64 pixelで  $\theta_n = \pi n/2$  とした。さらに、式(27)の閾値  $\xi$  は6とした。加えて、幅約241 mm、高さ約238 mm、奥行き約187 mmの物体を使用し、カメラから奥行き方向約1000 mmの位置で計測を行った。計測対象を図3の右に示す。一方、シミュレーションでは、Blender (Version 4.1.1) を用いて、

表1 3次元形状の平均距離誤差。  
Table 1 The mean distance error of 3D shapes.

手法	平均距離誤差 (mm)
従来手法 [3]	1.15
変位場 $\mathbf{h} = \mathbf{0}$ とした提案手法	0.76
完全な提案手法	<b>0.58</b>

これらの環境を再現した。

### 6.2 結果

全ての結果において、変位場の可視化では、変位場の  $xyz$  成分を色空間のRGB成分に対応づけ、変位場の  $-8$  mm から  $8$  mm の範囲を色空間の0から255の範囲に変換したものを示す。まず実環境で、図4の視線に対して、奥行き方向に約1000 mm/sで運動させた物体を提案手法で計測した。3枚の撮像画像と計測結果を図4に示す。並進運動を捉えながら、3枚の位相シフト法のパターンを用いて3次元形状を計測できたことが確認できた。

次に、シミュレーション上で、図5の視線に対して、右下奥方向に1732 mm/sで運動させた物体を計測した。この条件では、撮像画像の各画素に対応する光輸送係数がパターン間で大きく変化する。式(2)に基づく従来手法 [3]、変位場  $\mathbf{h} = \mathbf{0}$  とした提案手法、完全な提案手法でそれぞれ計測した結果を図5に示す。また、計測された3次元形状の平均距離誤差を表1に示す。提案手法は従来手法に比べて、50%精度が向上したことが確認できた。

さらに、シミュレーション上で、時計回りに毎秒3回転の速度で運動させた物体を提案手法で計測した。計測結果を図6に示す。回転運動を捉えながら、3次元形状を計測できたことが確認できた。

## 7 考察

図4, 5, 6に示すように、提案手法は物体の運動を捉え、3次元形状を計測することができた。しかし、図5は並進運動であるが、変位場にばらつきが見られる。変位場が空間的に滑らかになるような拘束を導入することで、さらに高精度化すると考えられる。

## 8 まとめ

本稿では、運動物体の高精度な3次元計測に向けた位相シフト法のためのニューラルインバースレンダリングを提案した。まず、運動物体の計測のために、物体の位置変化を表す変位場をレンダリングモデルに組み込み、物体の正確な動きを捉えた。さらに、運動物体を計測可能な位相シフト法に注目し、3枚のパターンで計測するために、カメラ複数台・プロジェクタ1台でのレンダリングモデルを構築した。

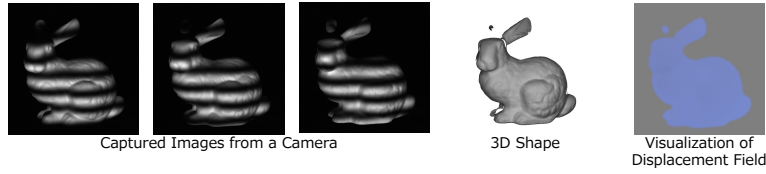


図4 3枚の撮像画像（左）と視線に対して奥行き方向に約 1000 mm/s で運動させた物体の計測（右）。

Fig. 4 Three captured images (left) and measurement of an object moving at approximately 1000 mm/s along the depth direction relative to the line of sight (right).

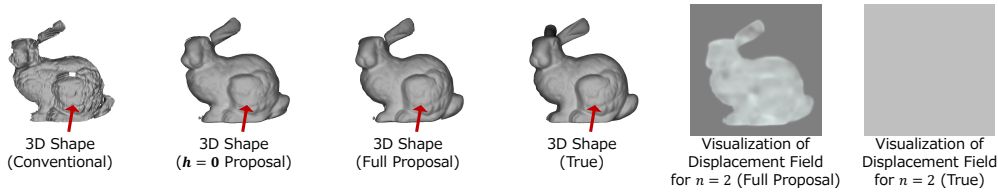


図5 視線に対して右下奥方向に 1732 mm/s で運動させた物体の計測。矢印部分のような凹凸に注目すると提案手法の有効性を確認できる。

Fig. 5 Measurement of an object moving at 1732 mm/s in the lower-right depth direction relative to the line of sight. By focusing on unevenness like that in the arrowed area, the effectiveness of the proposed method can be verified.

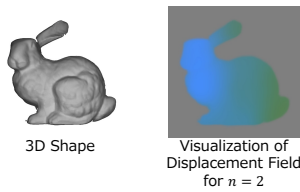


図6 時計回りに毎秒3回転の速度で運動させた物体の計測。

Fig. 6 Measurement of an object moving at a rotational speed of 3 revolutions per second in a clockwise direction.

実験では、カメラ2台・プロジェクタ1台の構成で3枚の位相シフト法のパターンを用いた運動物体の計測を達成したことを確認した。さらに、従来の運動物体を計測する位相シフト法と比較して、計測精度が50%向上したことを確認した。

#### 参考文献

- [1] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Advances in optics and photonics*, Vol. 3, No. 2, pp. 128–160, 2011.
- [2] Chao Zuo, Shijie Feng, Lei Huang, Tianyang Tao, Wei Yin, and Qian Chen. Phase shifting algorithms for fringe projection profilometry: A review. *Optics and Lasers in Engineering*, Vol. 109, pp. 23–59, 2018.
- [3] Thibaut Weise, Bastian Leibe, and Luc Van Gool. Fast 3D Scanning with Automatic Motion Compensation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2007.
- [4] Parsa Mirdehghan, Maxx Wu, Wenzheng Chen, David B. Lindell, and Kiriakos N. Kutulakos. TurboSL: Dense Accurate and Fast 3D by Neural Inverse Structured Light. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25067–25076, 2024.
- [5] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N. Kutulakos. Optimal Structured Light a la Carte. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6248–6257, 2018.
- [6] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10318–10327, 2021.
- [7] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*, Vol. 34, pp. 27171–27183, 2021.
- [8] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM transactions on graphics (TOG)*, Vol. 41, No. 4, pp. 1–15, 2022.
- [9] Yuan-Chen Guo. Instant Neural Surface Reconstruction. <https://github.com/bennyguo/instant-nsr-p1>, 2022.
- [10] Hwan Heo, Taekyung Kim, Jiyoun Lee, Jaewon Lee, Soohyun Kim, Hyunwoo J Kim, and Jin-Hwa Kim. Robust camera pose refinement for multi-resolution hash encoding. In *International Conference on Machine Learning (ICML)*, pp. 13000–13016, 2023.
- [11] Yoshihiro Watanabe and Masatoshi Ishikawa. High-speed and high-brightness color single-chip DLP projector using high-power LED-based light sources. In *International Display Workshops (IDW)*, pp. 1350–1352, 2019.