

深層学習によるプレーンスweep法の改善に関する 基礎研究

最上 達春(藤本・中谷研究室)

1. はじめに

プレーンスweep法は、複数のビデオカメラ映像から対象物体の表面の位置を推定することにより、リアルタイムで自由視点映像とデプス映像を生成する。従来の方法では、特に、シーン中に色の変化が少ない部分（テクスチャレス領域）、見る方向によって光の反射により輝度が異なる部分（ノンランバート領域）があるような場合には、生成される自由視点映像に乱れが生じ、高品質な映像が得られないことが多い。そこで、本研究では、深層学習を用いることでプレーンスweep法による生成画像の高品質化を目指す。

2. 関連技術

2.1. 深層学習

深層学習の基本的な仕組みを図1に示す。学習データとして、入力データと正解データを用意する。そして、入力データに予測計算を行い、出力データを得る。この出力データと正解データの損失を計算し、この損失を小さくするように勾配計算を行い、勾配値を出す。そして、勾配値をもとに予測関数のパラメータの修正を行い、再び入力データに予測計算を行う。この処理を繰り返すことでパラメータの最適化を目指すのが深層学習の基本的な仕組みである。

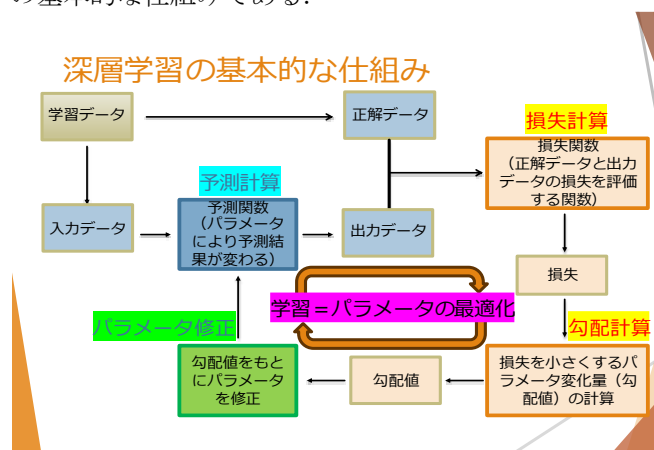


図1 深層学習の基本的な仕組み

2.2 プレーンスweep法

プレーンスweep法は、複数のカメラで撮影した画像を用いて自由視点から見た物体表面の形状と色を推定する手法である。図2のように、自由視点のスクリーンに平行な複数のプレーンを奥行き方向に等間隔に配置する。そして、各プレーンに各カメラの画像を投影し、自由視点画像の各画素に対応するプレーン上のサンプル点に投影されたカメラ画像の画素色（カメラ画素色）の類似性を求める。すべてのプレーンのうち、カメラ画素色の類似性が高く、同じ画素色とみなせるサンプル点を物体表面点と判定することで、自由視点から見える物体の形状（デプス）を推定する。そして、そのカメラ画素色から自由視点画像の画素色を得る。

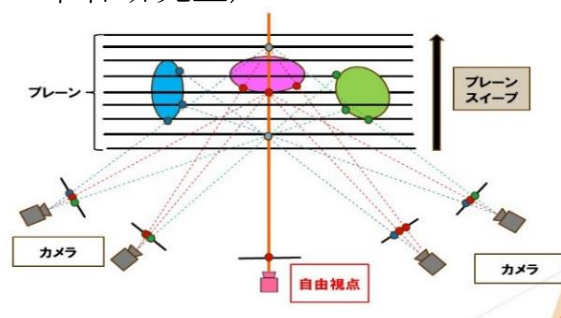


図2 プレーンスweep法の原理

3. 提案手法

3.1. データセットの作成

図3のように、物体を取り囲むようにカメラを配置する。円周方向の90度を9台で等間隔にカバーし、これを高さ方向に3段重ねることで計27台を配置する。そして、自由視点をカメラの位置にしてプレーンスweep法で生成された自由視点画像を学習用の入力データ、カメラで撮影された画像を正解データとし、深層学習のネットワークで入力画像から生成される出力画像が正解画像に近づくように、ネットワークの学習を行う。また、カメラがない位置に自由視点を置いたプレーンスweep法による自由視点画像を検証データとする。将来的には実空間で実施することを念頭に置き、本研究では仮想空間上で仮想モデルを用いて実施する。カメラの位置から仮想モデルをレンダリングした画像をカメラ画像とし、プレーンスweep法による自由視点画像を生成する。これにより、カメラがない位置を自由視点としたプレーンスweep法による自由視点画像に対しても、その自由視点に対して仮想モデルをレンダリングした正解画像により検証することができる。検証データでは、学習データによって訓練されたネットワークで入力画像から生成された出力画像がどれほど正解画像に近いかを検証するためだけに正解画像を用いる。

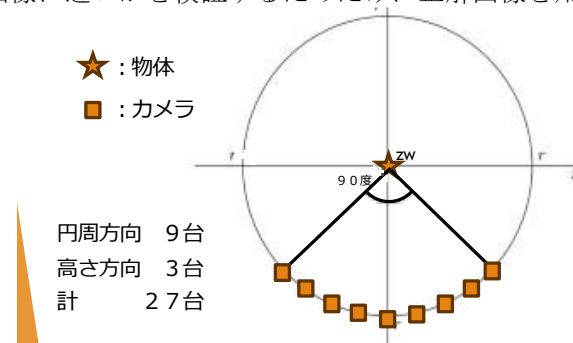


図3 カメラと物体の配置

3.2. 深層学習モデルの設計

予測計算を行うネットワークとして多層の畳み込み層を持つCNN（畳み込みニューラルネットワーク）モデルを用いる。各畳み込み層には64チャネルのフィルタを使用し、カーネルサイズを5×5、パディングを2とする。各畳み込み層の出力にはReLU関数を適用して非線形性を導入する。本研究実験では、畳み込み層の数を

変えて実験を行う。

3.3. モデルの学習と検証

本研究では、学習に用いる損失関数として2つの画像を比較する SSIM（構造的類似性指数）を使用する.SSIM は人間の視覚的特性を考慮した類似性の評価を行う.SSIM は0から1の値を取り、1の場合には2つの画像が同一であり,0の場合には全く類似性がない.勾配計算には誤差逆伝播法を使用する.また、パラメータを修正するオプティマイザとしては Adam オプティマイザを使用する.

また、検証には、SSIMに加えて PSNR も用いる. PSNR は MSE（平均2乗誤差）に基づいて計算され、次のように値が高いほど画像品質が高い.

30db：非常に高品質である.

20~30db：高品質ではあるが、わずかな劣化が目視できる.

20db 未満：明らかに劣化が見られる.

4.実験

仮想モデルとして,実験1ではティーポット,実験2では大小のさまざまな球を用い、畳み込み層の数を変えて実験を行った.その結果を図4、図5に示す.これらの結果は検証データに関するものである.入力画像はプレーンスweep法により生成された画像であり、出力画像は学習済みのネットワークにより入力画像から生成された画像である.学習率は 0.00005 とし、エポック数は 50 とした.また、学習前後の SSIM と PSNR を表1に示す.

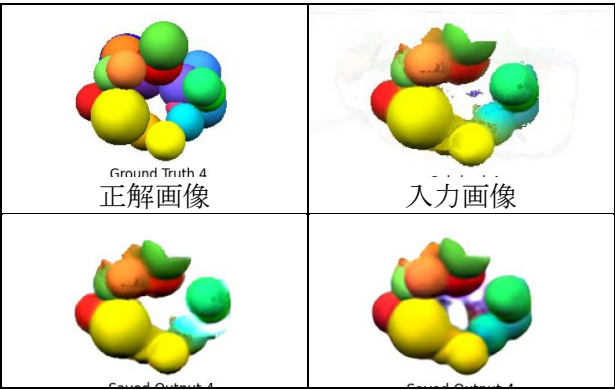
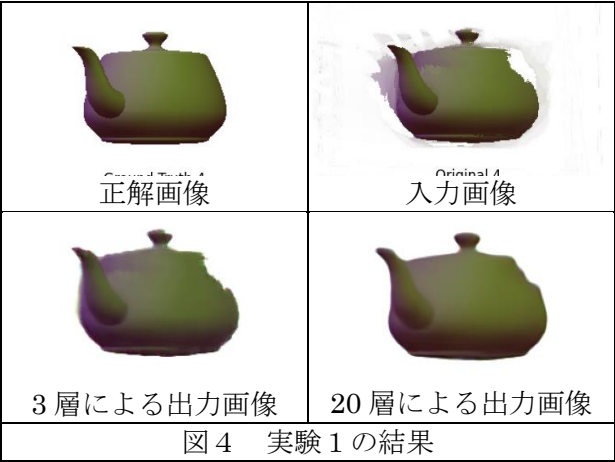


表1 学習前後の SSIM と PSNR の比較

	学習前 SSIM	学習後 SSIM	学習前 PSNR	学習後 PSNR
実験 1 20 層	0.8956	0.9648	20.26db	25.61db
実験 2 15 層	0.8681	0.9098	18.17db	19.22db

実験の結果から、深層学習を用いてプレーンスweep法で生成された画像をより精度の良いものにすることが確認された.畳み込み層の数が多いほど精度の良いものになった.これは、図4と図5に載せていない他の検証データについても同様であった.一方、実験2の20層では,モデル全体が黄色となり、学習がうまくできていない.この原因としては,層の数が増えるほど,連鎖的な計算の回数が増えるため、勾配の爆発が起こってしまった可能性が考えられる.

5.おわりに

本研究では、プレーンスweep法で生成された画像の精度を深層学習で向上させる方法について提案した.深層学習のネットワークで出力された画像は、目視だけではなく、SSIM や PSNR などの評価値もよくなっていることから、画像は改善されたといえる.今後の課題としては、学習率や損失関数を調整することで、より効率的に最適なネットワークを構築することが挙げられる.また、勾配の発散が起これないように残渣接続という方法をモデルに導入していきたい.そして、仮想空間ではなく実空間での実施、さらには、あらゆる物体やシーンに対して汎用的に適用可能であるネットワークの構築を考えていきたい.

参考文献

R. T. Collins, A Space-sweep Approach to True Multi-image Matching, Proc. of Conf. on Computer Vision and Pattern Recognition 1996, pp. 358-363, 1996.

Y. Liu, G. Chen, N. Max, C. Hofsetz, P. McGuinness, Undersampled Light Field Rendering by a Plane Sweep, Computer Graphics Forum, vol.25, no.2, pp.225-236, 2006.