

1. はじめに

物体の姿勢推定の手法の一つに、Inverting Neural Radiance Fields (iNeRF) [1] がある。iNeRF は、RGB 画像からカメラの姿勢推定が可能である。iNeRF に入力する画素は、中心サンプリングやランダムサンプリングにより決定する。これらのサンプリングでは、余分な背景情報が含まれたり、物体の一部のみが対象となることで推定誤差が生じることがある。正確に姿勢推定するためには、対象物体の領域を抽出するサンプリングが必要である。そこで本研究では、二値化処理により検出した物体範囲内から画素を決定する物体領域サンプリングを提案する。

2. iNeRF による姿勢推定

iNeRF は初期値のカメラ姿勢から目標画像のカメラ姿勢を推定する手法である。このとき、中心サンプリングまたはランダムサンプリングによって入力の画素を決定する。そして、カメラ姿勢から見た画素に対して仮想的な光線を投射し、この光線情報を pixelNeRF [2] に入力する。pixelNeRF の出力は、光線上の色と密度であり、ボリュームレンダリングにより光線に対する画素の画素値を決定する。レンダリングした画素と目標画像の画素の MSE を取り、誤差逆伝播法を使用してカメラ姿勢を更新する。

iNeRF のサンプリングにおいて、中心サンプリングは局所的、ランダムサンプリングは大域的であるため、余分な背景や物体の一部のみをサンプリングしてしまうことがある。そのため、姿勢推定に有効な情報を持つ物体の画素が選択されにくく推定精度が低下するという問題がある。

3. 提案手法

姿勢推定に有効な情報を持つ画素をサンプリングするため、対象物体を考慮した物体領域サンプリングを提案する。対象物体を検出するために、二値化処理を施す。物体領域サンプリングでは、二値化により検出した物体の範囲内からランダムに画素を決定する。各サンプリングの比較を図 1 に示す。このサンプリング方法を iNeRF に導入したネットワーク構造を図 2 に示す。まず初期画像と目標画像から物体の画素を検出する。次に、この 2 つの検出結果を結合し、物体の画素をサンプリング領域とする。このサンプリング領域の中からランダムに画素をサンプリングする。提案手法ではサンプリング方法のみを変更し、他は iNeRF と同様の処理を行う。2 回目以降のサンプリングは、更新されたカメラ姿勢からレンダリングされた画像と目標画像を二値化処理してサンプリング領域とする。



図 1：各サンプリングの比較

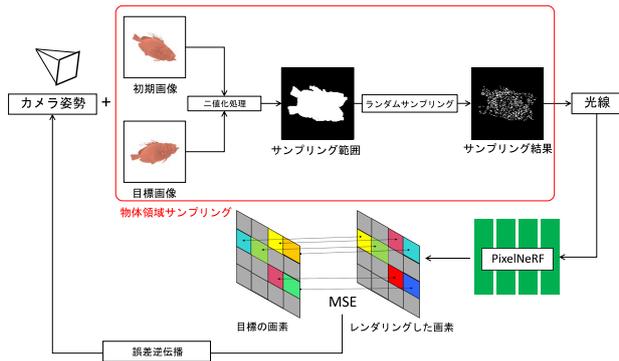


図 2：物体領域サンプリングを導入した iNeRF

4. 評価実験

本実験では、中心、ランダム、物体領域サンプリングを用いた iNeRF の姿勢推定精度を比較することで、提案手法の有効性を示す。

4.1. 実験条件

本実験では、サンドイッチ、マトウダイ、車、椅子のデータセットを [5,30] 度の範囲でランダムに回転させた 125 組用いてサンプリング方法を比較する。推論時の設定は、最適化手法を Adam、更新回数を 200、サンプリング数を 1024、損失関数を MSE とする。評価方法は、推定角度 $[x, y, z]$ の誤差平均を計算する。ここで、推定角度が 20 度以上の場合はエラーとする。

4.2. 実験結果

サンドイッチ、マトウダイ、車、椅子のデータセットを各サンプリングで姿勢推定した結果を表 1 に示す。() 内の数値は標準偏差である。表 1 より、物体領域サンプリングを用いた姿勢推定の精度が向上している。

椅子データを対象とし、更新回数 (step) が 0, 100, 200 のときの生成結果を図 3 に示す。図 3 より、物体領域サンプリングは更新回数が 100 回目のとき、目標画像に近い形状になっており、収束が速いことが確認できる。これにより、物体領域サンプリングを用いた iNeRF の有効性を示すことができた。

表 1：姿勢推定精度の比較

データセット	評価	中心	ランダム	物体領域
サンドイッチ	推定の誤差	4.80 (3.34)	4.94 (3.18)	4.51 (3.12)
	エラー率	0.16	0.14	0.11
マトウダイ	推定の誤差	6.03 (3.57)	6.03 (3.73)	5.69 (3.34)
	エラー率	0.39	0.04	0.24
車	推定の誤差	3.29 (2.80)	3.34 (2.61)	2.89 (2.65)
	エラー率	0.01	0.04	0.01
椅子	推定の誤差	4.10 (3.55)	4.12 (2.97)	2.83 (2.69)
	エラー率	0.16	0.15	0.12

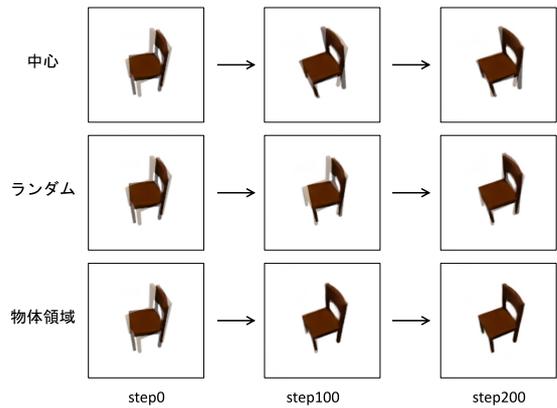


図 3：姿勢推定の可視化

5. おわりに

本研究では、二値化処理により検出した物体内からランダムに画素を決定する物体領域サンプリングを提案した。提案手法は、中心、ランダムサンプリングよりも姿勢推定精度が向上することを確認した。今後は、iNeRF の損失関数や最適化手法などを変更し、更なる精度向上を目指す。

参考文献

[1] L. Chen, *et al.*, “iNeRF: Inverting Neural Radiance Fields for Pose Estimation”, IROS, 2021.
 [2] A. Yu, *et al.*, “pixelnerf: Neural Radiance Fields from One or Few Image”, CVPR, 2021.