

単眼深度推定における物体認識の有効性

吾妻 福央[†] 平原 誠[†]

[†] 法政大学大学院 理工学研究科応用情報工学専攻 [†]

1. はじめに

単眼深度推定とは単眼カメラからの情報を基に深度(奥行き)を推定する技術[1]のことで、主に RGB 画像の入力から深度を推定する。本研究では物体認識画像をもモデルに追加入力することでその有効性を確認する。

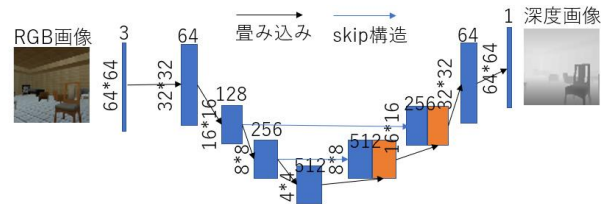


図1. 従来モデル[2]

2. 物体認識画像の追加意義

人間の単眼での奥行き知覚の手がかりとして像(網膜像)の大きさがある。我々は既知の物体とその像の大きさから深度を予測することが可能である。このように物体とその像の大きさとの関係を認識させるために、本研究では物体の種類ごとの物体認識画像をモデルに追加入力し深度推定精度の向上を図る。

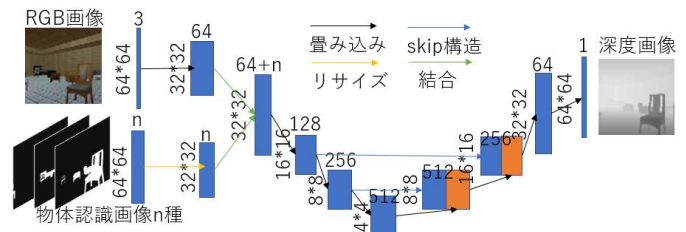


図2. 提案モデル

3. 従来モデルと提案モデル

以前我々は図1のように RGB 画像のみの入力から深度画像を出力するエンコーダデコーダ型モデル[2]を作成した。それに対して、今回提案するモデルは図2のように図1のモデルに物体認識画像を追加入力するようになっている。畳み込み層の活性化関数には Relu 関数, 出力層には Sigmoid 関数を用いる。誤差関数には誤差二乗和, フィルタの最適化には Adam を用い, ミニバッチ数8として学習させる。

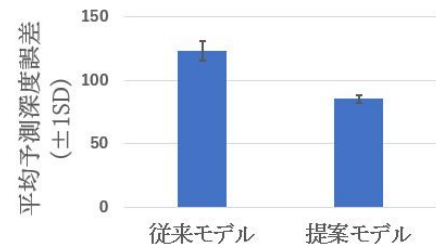


図3. 平均予測深度誤差(±1SD)の比較

4. データセット

本実験では、現実空間風に作成した部屋の VR データセットを用いる。部屋の中にはイスやテーブルなど7種類の家具を複数個設置した。7種類の家具は Segnet[3]のような物体認識モデルによって物体認識画像が作成できることを前提としており、今回は VR 空間で撮影した誤りのない物体認識画像を使用する。

計11部屋作成し、1部屋ごとに200セット撮影したため、計2200セットとなる。その内10部屋の1400セットを学習データ、600セットを検証データ、残り1部屋200セットをテストデータとした。

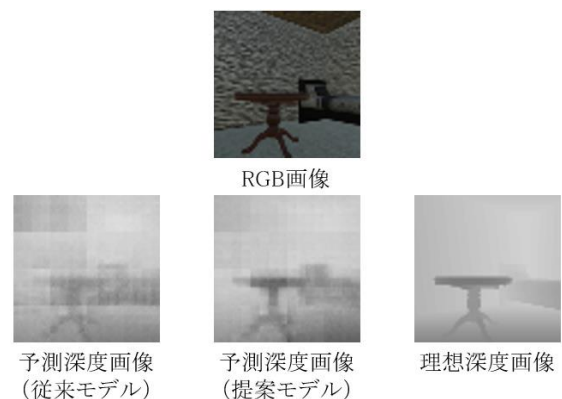


図4. 予測深度画像の比較

5. 実験結果とまとめ

テストデータに対する画像1枚あたりの平均予測深度誤差(±1SD)を図3に示す。平均予測深度誤差は提案モデルのほうが小さくなった。また、テストデータに対する予測深度画像の一例を図4に示す。提案モデルのほうが物体の深度と形をより正確に捉えることが出来た。これは物体認識画像によって物体の形状を把握し、その物体の像の大きさから深度を推定したためだと考えられる。

以上の結果より単眼深度推定において、物体認識画像は有効な入力であると言えるだろう。

参考文献

- [1] D.Eigen, *et al.*, NIPS14, Vol2, pp.2366-2374, 2014.
- [2] 吾妻,平原, SIC2020-2, pp.15-18, 2020.
- [3] V.Badrinarayanan, *et al.*, IEEE Trans. PAMI, Vol39, No.12, pp.2481-2495, 2017.