

単眼 RGB 画像からの深度画像生成モデルにおける前処理の検討

—暗所での精度に着目して—

西塚 諒一郎

1. はじめに

深度推定は、スマートフォンなどのデバイスから物体までの距離を測定したい場面を含め、幅広い分野で活用されている技術である。その中でも単眼深度推定は、従来のステレオカメラ¹やLiDAR² (Light Detection and Ranging)といった高価なセンサーを必要とせず、単眼カメラ³から取得した単眼RGB画像のみを用いて低コストで手軽な深度推定を実現する。

しかし、単眼深度推定にはいくつかの課題があり、その中でも撮影環境の影響を受けやすい点が問題視されている。特に暗所環境では、単眼深度推定の精度が著しく低下することが報告されており[1]、この問題の克服が強く求められる。暗所では単眼カメラにより撮影された画像の明度や彩度が低下するため、深度推定モデルが十分な特徴を抽出できないことが原因として考えられる。

本研究ではこの課題に着目し、単眼深度推定における前処理として、単眼RGB画像の明度および彩度の変化が深度推定の精度に与える影響を実験により示す。加えて、深度画像生成モデルの学習用データに施す最適な前処理を検討する。

2. 単眼深度推定の現状と課題

2.1 単眼深度推定について

単眼深度推定は、単眼カメラから得られたRGB画像に対応する深度情報を推定する技術である。この技術が活用されている分野は自動車またはドローンの自動運転[1]、拡張現実(AR)[2]、医療画像処理など多岐にわたる。

従来、深度推定を行うためにはステレオカメラ画像やLiDARの利用が必要とされてきた。しかし、これらのセンサーは比較的高価で、複数の機器を搭載するため重量が増すという問題があった[1]。

一方、機械学習技術と画像処理技術が近年大きく進歩した結果、単眼カメラから得られる情報だけで深度を推定することが可能になってきた。この技術はステレオカメラやLiDARを用いる場合に比べてコストや重量が大幅に抑えられ、より安価かつ軽量の深度推定システムの実現を可能にする。また、LiDARなどのセンサーをもともと搭載していないデバイスでも、単

眼カメラさえあれば深度推定が可能となるため、単眼深度推定の活用範囲をさらに広げられるという利点もある[3]。

2.2 単眼深度推定の課題

単眼深度推定には複数の課題があるが、とりわけ撮影条件に左右されやすく、ステレオカメラやLiDARに比べて推定精度が劣る点が大きな問題である。

例えば、撮影場所が暗所である、被写体に錯覚を引き起こすトリックアートが含まれる、光を強く反射する物体やテクスチャの少ない平坦面が写り込む場合などでは、ステレオカメラやLiDARを使用した場合に比べて精度が劣ることが報告されている[1]。

このように、単眼深度推定を用いた深度推定は、ステレオカメラやLiDARを使った方式よりも低コスト、軽量という点では勝っているが、深度推定の精度は劣るという課題がある。

2.3 単眼深度推定の現状

単眼深度推定法の研究では、近年、深層学習を用いた手法が注目を集めている。特に、CNN(畳み込みニューラルネットワーク)を基盤とするモデルが多くの研究で採用されており、画像のテクスチャや陰影、幾何学的な特徴を活用して深度情報を推定することが可能となった。

Eigenら[4]は、深度推定にCNNを初めて導入し、単眼カメラで取得した情報からの深度推定が可能であることを示した。この研究以降多くの改善案が提案され、効率と精度が大幅に向上している。

単眼画像からの深度推定における更なる精度向上を試みた例として、吾妻ら[6]の研究が挙げられる。単眼深度推定にCNNを使うという点では従来の手法と同じであるが、新たに物体認識画像を使用することで深度推定の精度向上を試みた。その結果、CNNに追加入力する画像として精度の高い物体認識画像を追加すれば深度推定の精度を向上できることを示した。一方で、認識させる物体の種類が増えるほどCNNへの入力画像が増え、学習や実装が煩雑になるという課題がある。

手代木[7]では深度推定を行うCNNのモデルの学習用データとして、深度情報に組み合わせる付加情報を検討した。その結果、深度推定における学習の

¹ ステレオカメラは2つのレンズから対象を測定する。

² レーザー光を照射し、その反射光を利用して距離を測定す

る技術。

³ 単眼カメラは1つのレンズから対象を測定する。

付加情報として、入力画像と対応する法線マップを使うことが最も効果的であり、精度向上の効果があることを示した。

しかし、これらの先行研究はいずれも比較的明るい場所での評価実験をしている。実験環境においては、暗所など撮影条件が厳しい環境でも高精度な深度推定が求められるが、暗所での推定精度を改善する具体的な手法を検討した研究は十分に行われていない。また、暗所を考慮するうえで重要となる画像の明度や彩度が深度推定に及ぼす影響についても、体系的に検証した先行研究は見当たらない。

そのため本研究では、RGB画像が持つ情報である明度と彩度が深度推定にもたらす影響に着目し、明度を低下させた画像を暗所の画像とみなして、単眼深度推定の暗所における精度向上を図る。

3. 実験に使用するシステムの概要

3.1 使用するツールとデータ

本研究で使用するシステムは、Keras⁴上で公開されているPython実装の深層学習ベースの深度推定ソースコードをもとに作成した。

深層学習に使用するデータセットとしてDIODE⁵データセットを使用する。DIODEデータセットはレーザーキャナーによって撮影されたデータをもとに作成されたデータセットである。それぞれ対応したRGB画像(.png)、深度画像(.npz)⁶、深度マスク画像(.npz)の3種類のデータで構成されている(図1)。モデルの学習と評価には正解深度画像(.npz)を使用する。正解深度画像(.npz)は深度画像(.npz)に対して、対応する深度マスク画像(.npz)を用いてマスク処理⁷を施すことで生成される(図2)。

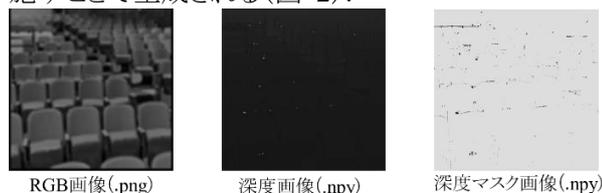


図1 実験に使用するデータの種類と拡張子

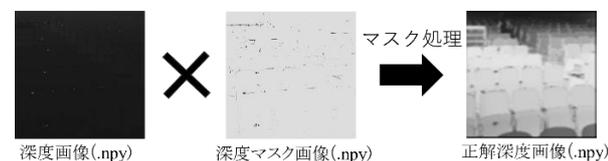


図2 正解深度画像の生成

3.2 彩度と明度について

本研究では使用するデータセットのRGB画像に対して、彩度と明度を変更したものを使用する。彩度と

明度は画像の特徴を表す重要な要素である。

彩度は色の鮮やかさや純度を示す指標である。具体的には、特定の色がどれだけ灰色(無彩色)を含まずに純粋であるかを表す。彩度が高い色は鮮明で鮮やかに見え、彩度が低い色は灰色がかかった色調に見える。彩度は色の純度や鮮やかさを評価する際の重要な指標となる[8]。

明度は色の明るさの度合いを示す指標であり、色がどれだけ白に近いか、または黒に近いかで決まる。これは光の反射量や透過量によって左右される。明度が高い色は明るく、低い色は暗く見える。明度は、色の明暗を評価する際の基本的な指標となる[8]。

本研究では画像の彩度と明度を低下させることによって暗所の画像を再現した(図3)。



図3 画像データにおける明度・彩度の比較

4. 予備実験

4.1 システムの概要

単眼RGB画像からの深度画像推定において、明度と彩度が深度推定の精度に及ぼす影響について予備実験で明らかにする。

システム構成については図4に示す。学習用データセットのRGB画像に対して、彩度もしくは明度を変更する前処理を施す点が他の研究と異なる特徴である。まず①で予備実験に使用するデータセットをDIODEのwebサイトからダウンロードする。次に②ではダウンロードしたデータのうち、学習用として260セットをコピーし、これをもう一組用意して合計520セットとする。③では2組あるデータセットの一方(260セット)のみに彩度または明度の変更を施す。残りの一方(260セット)はオリジナルのままの未処理の状態とする。この処理の結果、学習用データの内訳は前処理を施していないデータ260セットと、明度と彩度のどちらかを変更する処理を施したデータ260セットの合計520セットで構成される。なお、学習用データの彩度と明度を変更していない場合と比較するため、学習用データに明度または彩度の変更を施さなかった場合でも検証する(表1パターン①)。⑤ではこのデータを用いてモデル生成の深層学習を行う。最後の⑥において、学習データに使用していない別の65セットのデータを用意し、損失値によって精度を評価する。

⁴ https://keras.io/examples/vision/depth_estimation/.

⁵ <https://diode-dataset.org/>.

⁶ NumPy ライブラリによって提供されている形式。

⁷ 特定の部分のみを抽出し、それ以外を表示しなくする画像処理のこと。

予備実験は表 1に示した学習用データの7つのパターンについて、それぞれ30回⁸学習させる。本研究では深度推定を行う深層学習モデルの、学習用データに施す前処理に着目している。そのため学習回数に関しては、Kerasソースコードからの変更は行っていない。深層学習終了時に出力されたモデルの損失値をそのモデルの評価に使用する。ここでの損失値は、モデルが出力した深度画像と教師データである正解深度画像の差を算出したものであり、構造類似性指数、平均絶対誤差、深度平滑化損失を損失関数として使用している。損失値が0に近づくほど深度を正確に推定していることを意味する。

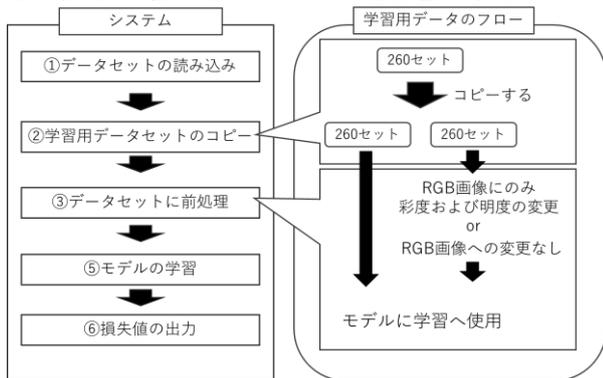


図 4 予備実験で使用したシステムの構成
表 1 学習用データの前処理パターン

	学習用データ	評価用データ
パターン①	オリジナル260枚+オリジナル260枚	オリジナル65枚
パターン②	オリジナル260枚+明度を減少させた260枚	オリジナル65枚
パターン③	オリジナル260枚+彩度を減少させた260枚	オリジナル65枚
パターン④	オリジナル260枚+明度を減少させた260枚	明度を減少させた65枚
パターン⑤	オリジナル260枚+彩度を減少させた260枚	明度を減少させた65枚
パターン⑥	オリジナル260枚+明度を減少させた260枚	彩度を減少させた65枚
パターン⑦	オリジナル260枚+彩度を減少させた260枚	彩度を減少させた65枚

4.2 予備実験の実験結果

予備実験により得られたモデルの損失値を整理し、その比較結果を表 2に示す。各行が学習用データの明度または彩度を何%減少させたかを示し、各列が評価用データの明度または彩度を何%減少させたかを示している。表中の各セルには、それぞれの低減率の組み合わせで深度推定を行った際の損失値を示した。表 2の項目のうち「オリジナル」は明度および彩度の値を変更していないデータを使用して

いることを表している。

この結果から、学習用データに明度減少画像を追加した場合、評価用データとして明度減少画像を用いた際の損失値がわずかに減少する傾向が確認された。このことは、学習段階で明度の減少に関する特徴を取り込むことにより、推定精度が若干ながら向上する可能性を示唆している。一方、学習用データに彩度減少画像を追加した場合は、明度減少画像での評価結果にほとんど差が見られなかった。これは、彩度の減少を考慮した学習が、明度の変化に対する性能改善には大きく寄与しないことを意味していると考えられる。

表 2 前処理タイプ別の精度比較

	オリジナル	評価用データ						
		明度25%減少	明度50%減少	明度75%減少	彩度25%減少	彩度50%減少	彩度75%減少	
学習用データ (前処理タイプ別)	オリジナル	0.5190	0.5603	0.6108	0.6762	0.5173	0.5146	0.5171
	明度25%減少	0.5210	0.5608	0.6091	0.6711	0.5193	0.5166	0.5146
	明度50%減少	0.5249	0.5626	0.6078	0.6652	0.5232	0.5206	0.5186
	明度75%減少	0.5320	0.5664	0.6068	0.6572	0.5304	0.5279	0.5261
	彩度0%減少	0.5190	0.5603	0.6108	0.6762	0.5173	0.5146	0.5171
	彩度25%減少	0.5235	0.5602	0.6107	0.6761	0.5171	0.5143	0.5121
	彩度50%減少	0.5187	0.5601	0.6107	0.6762	0.5169	0.5140	0.5118
	彩度75%減少	0.5235	0.5603	0.6108	0.6765	0.5169	0.5140	0.5117

さらに彩度を減少させた評価用データの損失値については、前処理を行っていない学習用データでも彩度減少の前処理をしても損失値がほとんど変化しないことが確認できる。この結果から、学習段階で彩度の変化を考慮する前処理を行う必要性は低いことが示唆される。したがって、深度推定精度の向上を目指す場合には、明度の減少に焦点を当てた学習用データの拡張がより効果的であると考えられる。

4.3 予備実験からの考察

学習用データに前処理を施すことにより、深度推定システムの損失値が減少することが確認できた。また、彩度を減少させた画像を学習用データとして用いる場合と比較して、明度を減少させた画像を使用したほうが、明度を減少させた評価用データに対してより効果的であることが示唆された。これらの知見から、学習用データとして明度減少画像を積極的に活用し、さらに学習用データのバリエーションと量を拡充することで、深度推定精度を一層高められると考えられる。本実験では特に明度に着目し、その効果を検証した。

5. 本実験

5.1 システム概要

本実験では、モデルの学習用データとして明度減少画像を使用する点に焦点を当てる。なお、予備実験のシステムをベースとし、学習用データの前処理部分におけるフローの一部を変更したシステムを用

⁸ Keras の標準設定である。

いる。変更後の学習用データのフローを図 5に示す。加えて本実験で使用した学習用データの前処理の組み合わせを表 3に示す。深度推定を行うモデルの学習用データとして、より多くのデータタイプを組み合わせたほうが深度推定への対応が柔軟となり、深度推定の精度向上に効果的だと考えた。そこで、前処理の複数パターンを組み合わせた学習用データを用意し、その精度を比較した。なお、モデルの評価には、予備実験と同じ損失値を指標として用いる。

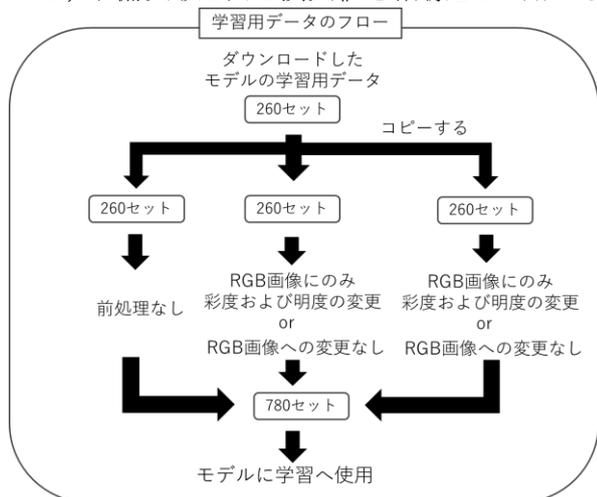


図 5 本実験における学習用データのフロー

表 3 本実験における学習用データの組み合わせ一覧

	学習用データ		
	タイプ1	タイプ2	タイプ3
パターン①	オリジナル	オリジナル	オリジナル
パターン②	オリジナル	オリジナル	明度減少
パターン③	オリジナル	明度減少	明度減少

5.2 実験結果と考察

本実験により得られたモデルの損失値を整理し、その比較結果を表 4に示す。表 4の第2列の項目名は、表 3におけるタイプ2とタイプ3の項目と対応しており、数値は明度減少の場合の減少率を示している。各学習用データと対応する評価用データにおけるモデルの損失値に加え、一番右の列に行ごとの損失値の平均値を算出した。

表 4 学習データの組み合わせ別の精度比較

		評価用データ				平均
		オリジナル	25%減少	50%減少	75%減少	
学習用データ (前処理タイプ別)	オリジナル+オリジナル	0.4624	0.5025	0.5539	0.6244	0.5358
	オリジナル+25%減少	0.4626	0.5017	0.5515	0.6198	0.5339
	オリジナル+50%減少	0.4642	0.5019	0.5494	0.6143	0.5325
	オリジナル+75%減少	0.4686	0.5040	0.5481	0.6074	0.5320
	25%減少+25%減少	0.4630	0.5012	0.5496	0.6159	0.5324
	25%減少+50%減少	0.4649	0.5017	0.5479	0.6018	0.5291
	25%減少+75%減少	0.4699	0.5044	0.5471	0.6056	0.5318
	50%減少+50%減少	0.4671	0.5026	0.5468	0.6068	0.5308
	50%減少+75%減少	0.4727	0.5060	0.5467	0.6013	0.5317
	75%減少+75%減少	0.4787	0.5100	0.5477	0.5975	0.5335

本研究の結果、学習用データに異なる明度の組み合わせを導入することで、各明度の評価用データにおける損失値がわずかに変化することが確認された。予備実験時と比較すると、特にオリジナル(明度を変更していない)状態の評価用データに対する損失値が、学習用データのいずれの組み合わせでも低減している。このことから、780セットまでであればモデルの学習用データを拡充することが、深度推定精度の向上に有効的であることが示された。また、損失値の平均値を比較したところ、学習用データとして「オリジナル+25%減少+50%減少」の組み合わせを用いた場合の損失値が0.5291となり最も低い損失値を示した。これらの結果から、暗所の画像と標準的な画像の双方に対してバランスよく深度推定を行うには、この組み合わせが最適である可能性が示唆された。

6. むすびに

本研究では、暗所での単眼深度推定の精度向上を目的として、画像の彩度または明度を調整する前処理の効果を検証した。その結果、学習用データとして「オリジナル+25%減少+50%減少」の組み合わせを用いた場合の損失値が0.5291となった。この結果より、暗所の画像と標準的な画像の双方に対してバランスよく深度推定を行うためには、この組み合わせが最適である可能性が示唆された。今後はより多様な種類のデータを用いて検証を行い、本手法の汎用性を確認する必要がある。

参考文献

- [1] 野波健蔵, 続・ドローン産業—応用のすべてで進化する自律飛行が変える未来—, 株式会社オーム社, 2020.
- [2] One Planet, AR オクルージョンとは, <https://1planet.co.jp/tech-blog/omRCqcvk>, (参照 2023-12-11).
- [3] 吉田顕・田村仁, “深層学習による深度画像推定を用いた AR アプリケーションの開発”, 情報処理学会第 81 回全国大会講演論文集, 2019 巻, 1 号, pp. 255 – 256, 2019.
- [4] EigenD, “Depthmap prediction from single image using a multi-scaled deep network” NIPS2014, pp. 2366–2374. 014.
- [5] 尾山匡浩, 松崎大輔, “単眼 RGB 画像からの深度画像生成”, ロボティック・メカトロニクス講演会講演概要集 2019, 2P2-H09-, 2019.
- [6] 吾妻福央, “単眼深度推定における物体認識の有効性”, 法政大学大学院紀要。理工学研究科編, 66 巻, pp. 1-8, 2022.
- [7] 手代木優哉, “画像生成モデルを用いた単一画像からの深度推定”, 東邦大学大学院理学研究科修士論文令和4年度, 2022.
- [8] 奥富正敏, デジタル画像処理[改定第 2 版], 画像情報教育振興協会, 2020.