Vol.2020-CVIM-222 No.18
2020/5/15

SLAMを用いたRoll方向回転に頑健な単眼depth推定の精度改善手法

齋藤　祐貴[1,a)]　八馬　遼[1,b)]　山口　真弘[1,c)]　斎藤　英雄[1,d)]

概要：本研究では，単眼 depth 推定において精度を改善する手法を提案する．近年，畳み込みニューラルネットワークを用いることにより，RGB 画像から精度の高い depth を推定することが可能になった．しかし，こうした従来研究では特にカメラが roll 方向回転をした際に depth の推定精度が低下する欠点がある．そこで本研究では RGB-SLAM のカメラトラッキング技術を適用し，カメラ姿勢を考慮した単眼 depth 推定の精度改善手法を提案する．本手法は RGB 画像のみを使用し，更にニューラルネットワークの構造変更や再学習の必要が無い．実験では，独自に作成したカメラの Roll 方向回転のみで構成されるデータセットと一般的な SLAM 評価用データセットを用いて，ベースライン手法に対する本手法の有効性を定性的，定量的に検証した．

In-Plane Rotation-Aware Monocular Depth Estimation using SLAM

1. はじめに

RGB 画像から物体までの距離 (depth) を推測する単眼 depth 推定は Augmented Reality (AR)[1] や自動運転 [2]，ロボットアプリケーション [3] 等に幅広く応用されている．近年，畳み込みニューラルネットワーク (CNN) を用いることで精度の高い depth 画像を推定可能になり [4], [5]，屋内，屋外シーンの学習用データセットも数多く登場してきた [6], [7].

CNN を用いた単眼 depth 推定では入力画像において，画像中に映る被写体の重力場方向がおおよそ真下を向いているという仮定が暗に存在する．これは人が写真を撮る際に，カメラを roll 方向に回転させないように撮るのが一般的であるためである．このため単眼 depth 推定では，depth を推定する際に入力画像中の重力場方向が強く推定精度に影響を与えると考えられる [8]．例えば，床や道路などカメラからの距離が比較的近くなる物体は画像中の下部に映り，天井や空などカメラからの距離が比較的遠くなる物体は画像中の上部に映る傾向がある．こうしたデータを用いて学習させることにより，画像下部は値が小さく，画像上部は値が大きい depth が推定される．

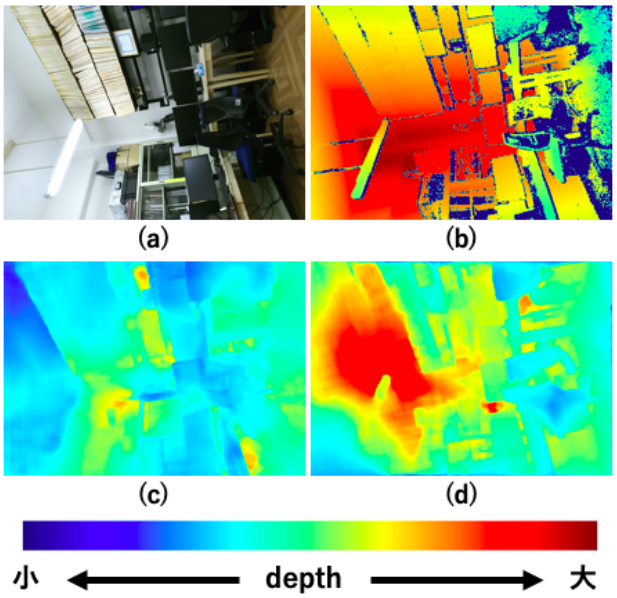

図 1　Roll 方向の回転を含むシーンの depth 推定結果．**(a)**: 入力画像, **(b)**: depth の真値, **(c)**: 従来手法の推定結果, **(d)**: 本手法の推定結果.

この重力場方向の仮定は，自動運転のようにカメラが固定されて，激しく動かないシーンが入力画像となる際には

[1] 慶應義塾大学
〒 223-8522 神奈川県横浜市港北区日吉 3-14-1
[a)] yusa19971015@keio.jp
[b)] ryo-hachiuma@keio.jp
[c)] yama-1467@keio.jp
[d)] hs@keio.jp

depth 推定の推定精度に大きな影響を与えない．しかし，スマートフォンのカメラやドローンカメラで撮影された画像を入力画像にする場面においては，ユーザーが自由にカメラ角度を操作可能なため，この重力場方向の仮定が容易に崩れてしまう．特にカメラが roll 方向回転をすることで，例えば距離が大きいはずの物体が画像下部に，距離が小さいはずの物体が画像上部に映る学習時と異なる位置関係が発生し，推定精度に大きな影響を与えると考えられる．

図 1 に具体例を示す．図 1(a) の画像は平行シーンを基準として，roll 方向に 90° 回転している．これをネットワークの入力画像とした際の推定結果が図 1(c) の depth 画像であり，床 (画像右部) と天井 (画像左部) が同じ depth を持つなど正しく推定されていないことが明らかである．

本論文では，CNN を用いた単眼 depth 推定の roll 方向回転によって推定精度が低下する欠点を RGB-SLAM を用いて精度向上させる手法を提案する．RGB-SLAM は RGB 画像シーケンスのみを用いて，xyz の平行移動成分と xyz 軸の回転成分から構成される 6 自由度のカメラ姿勢を推定する．カメラ姿勢を分解することで roll 方向の回転成分が抽出され，これを用いて入力画像にアフィン変換を適用し，カメラがあたかも roll 方向に回転していないような画像を生成する．この変換後の画像を depth 推定ネットワークに入力し，その出力結果に再度，逆方向の回転変換をするアフィン変換を適用することで，入力画像と同解像度の depth 画像が推定される．図 1(d) の depth 画像が本手法の推定結果であり，図 1(c) の結果と比べて図 1(b) の真値に近いより高精度な推定ができていることが分かる．

実験では，公開されている SLAM 評価用の TUM RGB-D データセット [9] と独自に作成した roll 方向のカメラ運動のみで構成される RGB-D データセットの両方を用いて評価を行った．TUM RGB-D データセットはカメラが pitch, yaw, roll の順で回転していく 3 つの *rpy* シーケンスを用いた．また，独自に作成したデータセットは 6 つの屋内シーケンスで構成され，カメラは-180 $< \theta <$ 180 の roll 回転のみ行う．これら 2 つのデータセットを用い，ベースライン手法と比較することで本手法の有効性を定性的，定量的に考察した．

2. 関連研究

2.1 roll 方向回転を考慮した推定

Toyoda らは人間が側転など激しい運動をするビデオ映像から CNN を用いた骨格推定を行う際に，同一の入力画像を様々な角度で回転させ最も信頼値の高い結果を採用することで人間の骨格推定の精度を改善した [13]．被写体が上下逆になるようなシーンは実世界では稀であり，データセットにこうしたデータは含まれない．この手法によってデータを再学習させる時間や計算コストを抑えることが可能になった．しかし、同一の入力画像を roll 方向に一定間隔の角度で回転させていく手法は計算コストが高く，リアルタイム性が低い．そこで，本手法では RGB-SLAM のリアルタイムで計算されるカメラ姿勢を直接用いることで，より正確で計算コストの低いシステムを提案する．

また，Kurz らは慣性センサー (IMU) を用いて SLAM のカメラトラッキングの精度を向上させる手法を提案した [14]．これは重力場方向を考慮して画像を用いた 3 次元センシングの精度向上を図った点で本研究と類似している．しかし，本手法では慣性センサーや depth センサーなど高機能センサーを一切用いず，RGB 情報のみで動作する手法を提案する．

2.2 単眼 depth 推定を組み合わせた RGB-SLAM

RGB-SLAM は特徴点ベースの手法 [15], **?**や輝度値ベースの手法 [16], [17], [18] 等が提案されてきたが，これらの手法はテクスチャが十分なシーンにおいても回転運動をするカメラ姿勢を精度よく推定できない欠点がある．

そこで密な点群を復元し，テクスチャがあまりないシーンにおいてもカメラ姿勢を正しく推定するため，RGB-SLAM と単眼 depth 推定を組み合わせた手法が提案されている．代表例が CNN-SLAM [11] であり，CNN を用いて推定された depth が輝度ベースの RGB-SLAM に組み込まれている．CNN-SLAM は roll 方向回転が起こるシーンでも正しくカメラ姿勢と密な環境地図を作成可能だが，roll 方向回転時には KeyFrame が生成されず depth を推定できない．そのため CNN に roll 方向回転を含むシーンを入力すると，1 章で示したような精度低下が起こると予想される．また，CNN-MonoFusion[1] は TUM RGB-D データセット [9] の *rpy* シーケンスの 3 次元復元結果を評価しているが，CNN の推定 depth の精度を定量的に評価していない．本稿では，学習データが原因となる単眼 depth 推定の欠点に問題を設定し，本手法の有効性を定性的，定量的に評価する．

2.3 CNN を用いた roll 方向回転角度の推定

1 枚の RGB 画像のみからカメラの roll 方向回転角を推定する手法が提案されている．例えば，Fischer らは画像の roll 方向回転角を回帰問題により推論するネットワークを提案した [19]．Greg らは 1 枚の RGB 画像から pitch, roll 方向の回転角を推定する CNN を用いた手法を提案した [20]．しかし，これらの手法はおおよそのカメラ回転角しか求まらず精度が不十分である．また，幾何的拘束条件も考慮せず，学習データに基づいた推定しか行えない．

また，Xian らは局所的，大域的な特徴を 1 枚の RGB 画像から抽出することで 2 自由度のカメラ回転角を推定している [21]．しかし，この手法は ±20° もしくは ±50° のような小さな回転角でしか有効性を検証していない，

本手法では幾何情報を考慮した回転角の推定手法として ORB-SLAM2[22] を用いる．ORB-SLAM2 は高精度に

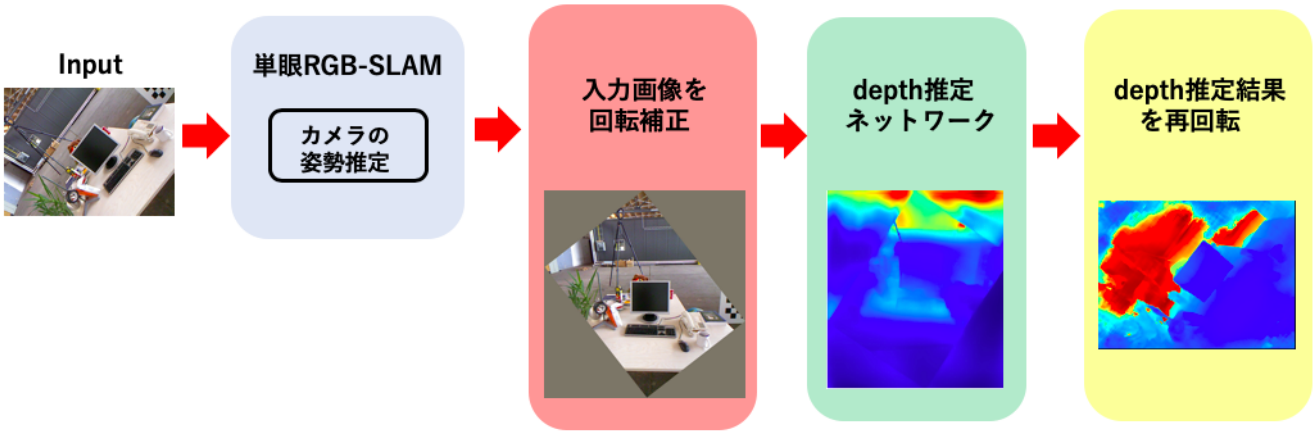

図 2　提案手法の概要図

カメラ姿勢を計算でき，比較的大きな roll 方向回転角も精度よく推定できる．

3. 提案手法

図 2 に本手法の概要図を示す．まず，はじめに RGB-SLAM を用いて入力 RGB 画像のカメラ姿勢を推定する．次に SLAM のカメラ姿勢から KeyFrame 画像を回転させるアフィン変換 $F(\theta)$ を計算し，あたかも roll 回転していないような画像を生成する．そして，この画像を CNN に入力する．最後に出力の depth 画像に対して先ほどとは逆向きの回転を適用することで入力画像と同解像度の depth 画像を推定できる．

3.1 RGB-SLAM を用いたカメラ姿勢推定

単眼 RGB-SLAM は多視点幾何を用いて正確なカメラ姿勢を推定できる．カメラが回転運動をした際にも高精度なカメラ姿勢を推定するために，本手法は ORB-SLAM2 [22] を用いた．輝度ベースの RGB-SLAM[16], [17] に比べ，ORB-SLAM は計算コストが低い利点がある．

3.2 CNN を用いた単眼 depth 推定

depth 推定部には CNN-MonoFusion[1] と同じ構造のネットワークを用いた．このネットワークは ResNet50[23] に基づいており，異なる焦点距離を持つカメラで撮影された複数のデータセットで学習を行えるよう推定 depth を正規化する *AdaBerhu* 損失関数を採用している．学習時の内部パラメータと異なるカメラで撮影された入力画像の depth を絶対スケールで計算するために，ネットワークの出力 depth は以下の式 1 を用いて SLAM のスケールに変換される．

$$D_{test} = \frac{f_{test}}{f_{tr}} D_{CNN} \tag{1}$$

ここで，D_{CNN} はネットワークの出力 depth，f_{test} は SLAM に用いられるカメラの焦点距離，f_{tr} は学習時に用いられるデータの焦点距離である．

3.3 回転角を用いた画像補正

depth 推定ネットワークに入力する前にアフィン変換を RGB 画像に適用し，被写体が平行シーンになるよう画像を拡大，回転させる．この際 SLAM は初期フレームを基準としてカメラ姿勢を推定するため，今回 SLAM の初期フレームではカメラは一切 roll 方向回転をしていないと仮定する．

ORB-SLAM2 [22] を用いて推定されたカメラ姿勢 $\mathbf{T}_t^{cw} \in R^{4\times4}$ は回転成分 $\mathbf{R}_t^{cw} \in R^{3\times3}$ と平行移動成分 $\mathbf{s}_t^{cw} \in R^3$ から構成される．この内，カメラ回転成分 $\mathbf{R}^{cw}$ は式 2 のように xyz 3 軸周りの各成分へと分解できる．

$$\mathbf{R}^{cw} = \mathbf{R}^{cw}(\psi)\mathbf{R}^{cw}(\phi)\mathbf{R}^{cw}(\theta) \tag{2}$$

ここで，ψ は pitch, ϕ は yaw, そして θ は roll 方向の回転を表す．Roll 方向の回転に着目し，$\mathbf{R}^{cw}(\theta)$ は式 3 のような 3×3 の行列で表現される．

$$\mathbf{R}^{cw}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

この θ を用いて，式 4 に示す 2×3 の $F_t(\theta)$ アフィン変換を入力画像に適用する．

$$F_t(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & s_x \\ \sin\theta & \cos\theta & s_y \end{pmatrix} \tag{4}$$

s_x，s_y は元の RGB 画像の中心を変換後の画像の中心に平行移動させるベクトルである．

アフィン変換 $F_t(\theta)$ で画像を回転させる際にはバイリニア補間を適用し，元画像の pixel を完全に残し余白の pixel を最小限にするよう画像サイズを調整する．また，畳み込み演算の際に余白の pixel が影響を与えないようにするために，CNN の畳み込み層では余白の pixel 値を 0 に設定

する．

CNN の出力が得られたら，再度 depth 画像を逆方向に回転させるアフィン変換を適用し，元の RGB 画像と同解像度の depth 画像を得る．

4. 実験

4.1 実験の詳細

CNN に RGB 画像を直接入力するベースライン手法と比較して本手法を評価する．depth 推定のネットワークは本手法とベースライン手法で同じであるり，本手法は CNN の前後で画像をアフィン変換させる点のみが両者の違いである．実験環境は Intel Core i7-7700 CPU と Nvidia GTX 1080Ti GPU 搭載のデスクトップ PC を用い，定性的，定量的評価を行った．

depth 推定ネットワークでは CNN-MonoFusion[1] からオンライン上で公開されている既存の学習済みモデルを用いた[*1]．このモデルは NEAIR データセット [1] と NYU Depth V2 データセット [6] の両方を用いて学習されており．ネットワークの出力 depth は入力画像の縦横半分の解像度なので，元画像と同じ解像度へと拡大した．

4.2 評価に用いたデータセット

まず，Kinect V1 で撮影され，カメラが回転運動する TUM RGB-D データセット [9] の 3*rpy* シーケンスを用いた．Kienct V1 は 4.0m 以上の depth を取得できないため，TUM データセットは真値の depth が 0 である pixel を多数含んでいる．また，roll 方向の回転角によって真値のフレーム枚数にばらつきがある．例えば，90˚ 付近の真値のフレーム数は 0˚ 付近のフレーム枚数に比べて極端に少ない．

以上 2 点の理由から，Kinect V2 を用いて自作のデータセットを作成した．このデータセットは 6 シーケンスから構成され，$-180 < \theta < 180$ のフレーム枚数が均等になるような roll 方向回転を行う．各シーケンスは約 30，80 秒であり，解像度 640×480 の RGB 画像と真値の depth 画像 7704 枚を含んでいる．全シーケンスは屋内環境で撮影され、得られる Depth の最大値は、Kinectv2 の取得可能範囲である、4.5m 以内となっている。

5. 評価

5.1 定性的評価

図 3 に TUM RGB-D データセットと自作データセットの定性的結果を示す．ベースライン手法に比べて本手法の方が真値に近く，depth の精度が改善されていることが分かる．例えば，図 3 の seq1 $\theta = 125$ では本手法は天井を机部分より遠くに推定しているが，ベースライン手法は天井が机や椅子の領域と同じ depth であると推定している．

[*1] https://github.com/NetEaseAI-CVLab/CNN-MonoFusion

表 1 自作データセットの平均誤差

	Abs_Rel ↓		RMSE ↓	
	本手法	ベースライン	本手法	ベースライン
seq1	**0.2372**	0.3260	**0.7865**	1.0565
seq2	**0.5161**	0.5929	**0.6557**	0.7321
seq3	**0.2745**	0.3425	**0.9406**	1.2068
seq4	**0.3590**	0.4225	**1.2818**	1.4601
seq5	**0.3687**	0.3930	1.9395	**1.9262**
seq6	**0.2614**	0.3339	**0.8769**	1.1095
ave	**0.3169**	0.3818	**1.0961**	1.2784

5.2 定量的評価

今回評価には絶対相対誤差 (Abs_Rel) と平均平方二乗誤差 (RMSE) の 2 つの指標を用いた．値がより小さい手法の方が精度が高いと言える．まず表 1 に自作データセットの平均結果を示す．AbsRel, RMSE 共に本手法はベースライン手法の精度を上回っていることが分かる．また図 5 に自作データセットのシーケンスごとの結果 (Abs_Rel) を示す．横軸は roll 方向の回転角を示し，縦軸は誤差値を示す．$\theta = 0$ 付近の誤差値は本手法とベースライン手法で大差がないが，$-45 < \theta < 45$ 付近でベースライン手法の精度が本手法よりやや高いシーケンスがある．これは 3 節で述べたように，本手法で画像をアフィン変換した際の余分なピクセルが影響を与えたと考えられる．しかし，$-\theta < -45$，$45 < \theta$ ではベースライン手法の誤差値が大きくなる一方で，提案手法の誤差値にあまり変化がなく一定となっている．roll 方向の回転角が大きくなった際に本手法の有効性を確認できた．

次に表 2 に TUM データセットの平均結果を，図 4 にシーケンスごとの結果 (Abs_Rel) を示す．TUM データセットではベースライン手法が本手法の精度を上回っていることが分かる．しかし，図 3 で示した TUM frei1_rpy の定性的評価では本手法の方がより真値に近い depth を推定できている．

この定量的評価で精度が上回らなかった原因は 2 つあると考えられる． 1 つ目は TUM データセットの frei2_rpy，frei3_rpy では Kinect V1 の depth 取得範囲を超えた depth を持つシーンがあり，depth の真値が 0 であるデータを多数含んでいるためである．これらの pixel 値は定量的評価で評価に含めないため，提案手法の有効性が確認できなかったと考えられる． 2 つ目は TUM データセットではカメラの roll 方向回転角が比較的大きい際，フレーム枚数が少ない傾向があるためである．フレーム枚数にばらつきがある中で平均値を比較したことで公平な評価ができなかったと考えられる．

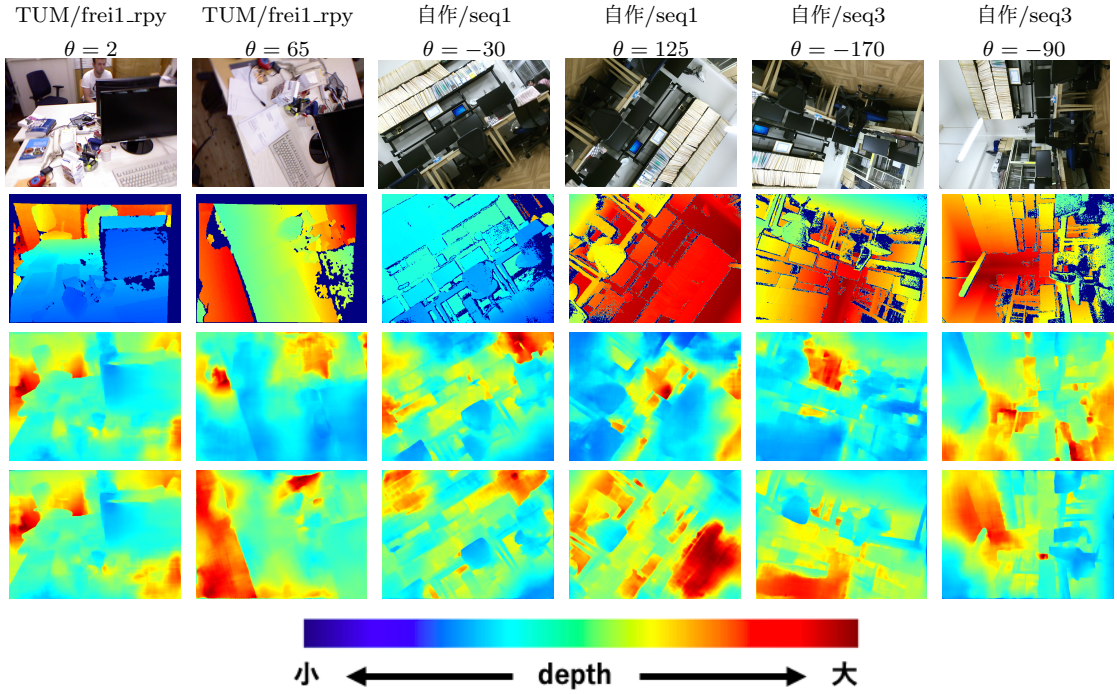

図 3 TUM データセットと自作データセットの定性的評価結果．上から順に入力画像，depth の真値，ベースライン手法の推定結果，本手法の推定結果．

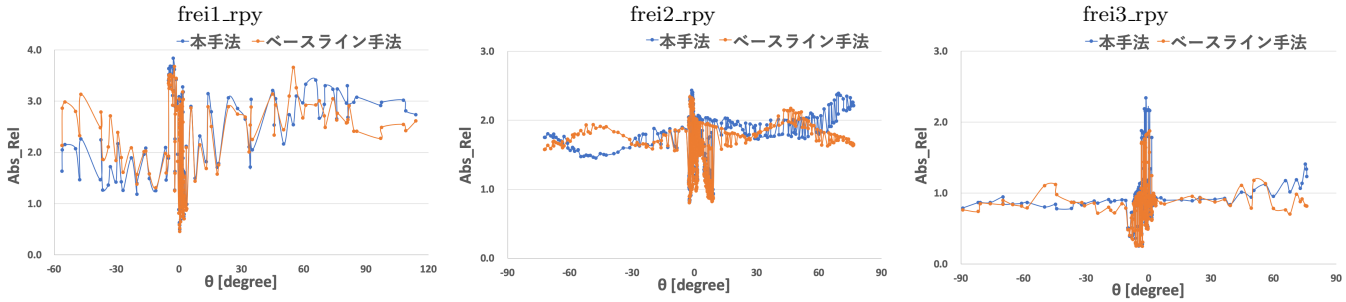

図 4 TUM データセットにおける roll 方向回転角と Abs_Rel 値の関係

表 2 TUM データセットのの平均誤差

	Abs_Rel ↓		RMSE ↓	
	本手法	ベースライン	本手法	ベースライン
frei1_rpy	2.2482	**2.2380**	**1.0036**	1.0217
frei2_rpy	1.7147	**1.6586**	0.9486	**0.9285**
frei3_rpy	0.9190	**0.8557**	0.9062	**0.8672**
ave	1.6211	**1.5708**	0.9480	**0.9297**

6. おわりに

本研究ではカメラ姿勢を直接用いることによって roll 方向回転が起こるシーンで単眼 depth 推定の精度を改善する手法を提案した．CNN の前後で画像にアフィン変換を加える単純な手法だが，一般的なデータセットと自作データセットで評価することで本手法の有効性を定性的，定量的に確認できた．本手法はカメラ姿勢の計算に RGB-SLAM を用いることで RGB 情報のみで動作可能であり，CNN の構造や学習に一切依存しない．

今後，RGB-SLAM を用いた際の初期フレームが roll 回転していないという仮定を取り除く予定である．また，本手法の精度を様々な角度でデータ拡張して学習させたモデルと比較する評価を検討したい．

参考文献

[1] Wang, J., Liu, H., Cong, L., Xiahou, Z., Wang, L.: CNN-MonoFusion: Online Monocular Dense Reconstruction Using Learned Depth from Single View. In: IEEE International Symposium on Mixed and Augmented Reality Adjunct(ISMAR-Adjunct), pp. 57-62. IEEE, Munich (2018).

[2] Wang, Y., Chao, W., Garg, D., Hariharan, B., Campbell, M., Weinberger, K.: Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving. In: In: Proceedings of the IEEE International Conference on Computer Vision, pp. 8445-8453, IEEE (2019)

[3] Marcu, A., Costea, D., Licaret, V., Pirvu, M., Slusanschi, E., Leordeanu, M.: SafeUAV: Learning to estimate depth and safe landing areas for UAVs from synthetic data. Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Work-

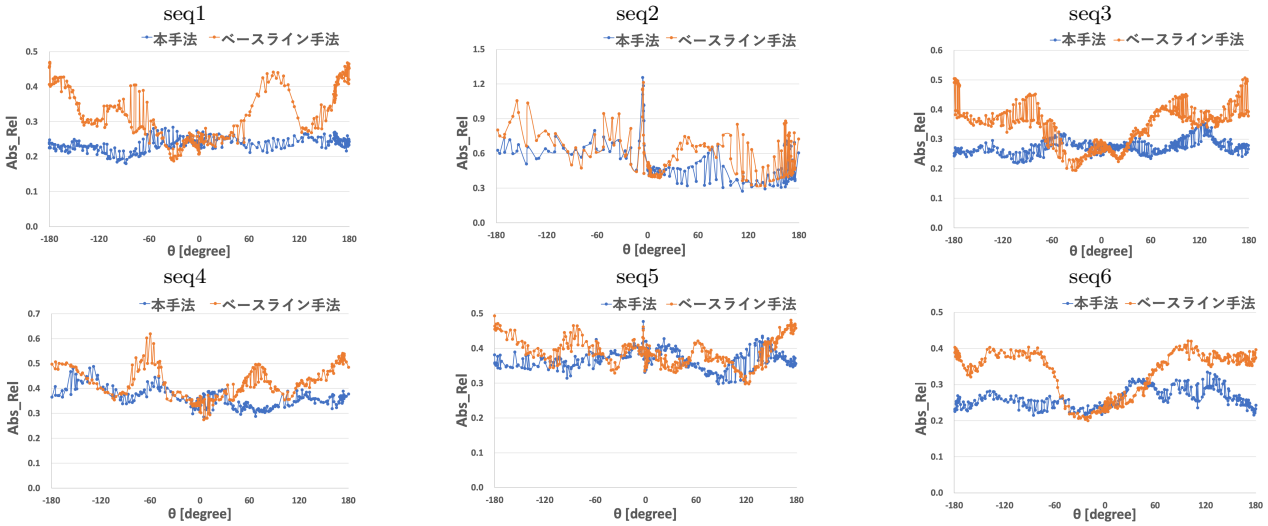

図 5　自作データセットにおける roll 方向回転角と Abs_Rel 値の関係

shops. LNCS, vol. 11130, pp. 43–58. Springer, Cham (2019).
[4] Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: International Conference on 3D Vision(3DV), pp. 11–20. IEEE (2016)
[5] Fu, H., Gong, M., Wang, C., Batmanghelich, K., Tao, D.: Deep Ordinal Regression Network for Monocular Depth Estimation. In: In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2002-2011. IEEE (2018)
[6] Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor Segmentation and Support Inference from RGBD Images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7576, pp. 746–760. Springer, Heidelberg (2012)
[7] Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A.: Sparsity invariant cnns. In: International Conference on 3D Vision (3DV), pp. 11–20. IEEE (2017)
[8] Mi, L., Wang, Hao., Tian, Y., Shavit, N.: Training-Free Uncertainty Estimation for Neural Networks. In: arXiv preprint arXiv:1910.04858 (2019)
[9] Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A Benchmark for the Evaluation of RGB-D SLAM Systems. In: IEEE International Conference on Intelligent Robot Systems, pp. 573–580. IEEE (2012)
[10] Grisettiyz, G., Stachniss, C., Burgard, W.: Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In: IEEE International Conference on Robotics and Automation, pp. 2432–2437. IEEE (2005)
[11] Tateno, K., Tombari, F., Laina, I., Navab, N.: CNN-SLAM: Real-Time Dense Monocular SLAM With Learned Depth Prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6243-6252. IEEE (2017)
[12] Laidlow, T., Czarnowski, J., Leutenegger, S.: DeepFusion: Real-Time Dense 3D Reconstruction for Monocular SLAM using Single-View Depth and Gradient Predictions. In: International Conference on Robotics and Automation, pp.4068–4074. IEEE (2019)
[13] Toyoda, K., Kono, M., Rekimoto, J.: Post-Data Augmentation to Improve Deep Pose Estimation of Extreme and Wild Motions. In: arXiv preprint arXiv:1902.04250 (2019)
[14] Kurz, D., Benhimane, S.: Gravity-aware handheld Augmented Reality. In: IEEE International Symposium on Mixed and Augmented Reality, pp 111-120. IEEE (2011)
[15] Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality, pp 0–10. IEEE (2007)
[16] Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-Scale Direct Monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part II. LNCS, vol. 8690, pp. 834–849. Springer, Heidelberg (2014)
[17] Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE transactions on pattern analysis and machine intelligence **40**(3), 611–625 (2017)
[18] Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: Fast semi-direct monocular visual odometry. In: International Conference on Robotics and Automation, pp.15–22. IEEE (2014)
[19] Fischer, P., Dosovitskiy, A., Brox, T.: Image orientation estimation with convolutional networks. In: Gall, J., Gehler, P., Leibe, B. (eds.) GCPR 2015. LNCS, vol. 9358, pp. 368–378. Springer, Cham (2015)
[20] Olmschenk, G., Tang, H., Zhu, Z.: Pitch and roll camera orientation from a single 2D image using convolutional neural networks. In: In 2017 14th Conference on Computer and Robot Vision, pp.261–268. IEEE (2015)
[21] Xian, W., Li, Z., Fisher, M., Eisenmann, J., Shechtman, E., Snavely, N.: UprightNet: Geometry-Aware Camera Orientation Estimation from Single Images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9974–9983. IEEE (2019)
[22] Mur-Artal, R., Tardós, J. D.: Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE transactions on robotics **33**(5), 1255–1262 (2015)
[23] He, K., Xiangyu Z., Shaoqing R., Jian, S.: Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 770-778. IEEE (2016)