車載動画像を用いた道路空間の深度判定の高精度化に向けたパラメータ補正

佐藤 剛*・前田 紘弥**・樫山 武浩***・関本 義秀**

Parameter Correction for Accurate Depth Estimation in Road Space Using In-vehicle Video

Go Sato*, Hiroya Maeda**, Takehiro Kahiyama***, Yoshihide Sekimoto**

Abstract: The measurement of the distance to the object in the image is called depth estimation. It can be used to estimate the 3D shape of the object in the image. In this paper, depth estimation from monocular images is investigated using deep learning. One of the challenges of depth estimation from monocular images is that existing studies only estimate relative distance, though absolute depth estimation is required for object size estimation. In this paper, we construct an absolute depth estimation model using monocular images. The internal parameters of existing models are adjusted using in-vehicle video taken in Saitama. The estimated relative depth is modified to the absolute depth using the tilt and height of the camera. The estimated depth is evaluated using stereo camera images.

Keywords: 車載動画像(in-vehicle video)，深層学習(deep learning)，深度推定(depth estimation)

１．はじめに

画像内の物体までの距離を成分として持つ画像を作ることを深度推定といい， 3D モデリングや拡張現実(AR)への応用が可能である．特に車載カメラ撮影画像の深度推定は，自動運転や道路維持管理に直接利用できるため，重要度は高い．そこで本論文では車載カメラ撮影画像からの深度推定を考える．

深度推定の既存手法としてはレーザースキャナやステレオカメラがある．だがこうした器具は高価であり，また得られる深度画像が疎であるという欠点を持つ．こうした欠点を克服すべく，単眼画像を用いた深度推定の研究が進められている．

単眼画像を用いた深度推定の手法として，車載動画像を用いて深度判定モデルを訓練し，実運用時は単眼画像から深度推定を行う手法が提案されている．こうしたモデルは訓練・運用共に比較的容易で，得られる深度画像も密であるという利点を持つ．

一方，このような深層学習ベースの深度推定モデルを日本国内で活用する際の問題点として，推定された深度は実際の深度に対し定数倍の誤差を持つ相対深度であるため，そのままでは自動運転や建設工事に応用することが難しい点が挙げられる.加えて，既存モデルは海外の車載動画像によって訓練されており，日本国内での利用時に精度が担保されない．

そこで本論文では，海外の車載動画像によって訓練した深度推定モデルのパラメータを埼玉県内の道路を撮影した車載動画像により調整し，相対深度判定精度が改善することを確認した．その上で，加速度センサから求めたカメラ傾斜角に基づく推定値に画像中央部の深度が一致するように相対深度推定結果を定数倍し，絶対深度を取得した．取得した絶対深度を，相対深度推定値をそのまま絶対深度として用いた場合と比較した結果，深度推定評価指標(Absolute Relative Error)が約 40％改善した．

*　学生会員　東京大学工学部社会基盤学科 (Department of Civil Engineering, the University of Tokyo)
Email: gosato@iis.u-tokyo.ac.jp
**　正会員　東京大学生産技術研究所 (Institute of Industrial Science, the University of Tokyo)
***　非会員　東京大学生産技術研究所 (Institute of Industrial Science, the University of Tokyo)

２．既存研究の紹介

単眼画像からの深度推定の方法として，動画像を用いて訓練した深層学習モデルを提案した研究にZhou et al. (2017) がある．一般に動画像の隣接フレームには共通物体が映る．時刻 T と時刻 T+ΔT を考えると，フレーム間のカメラの動きは回転行列 R と移動ベクトル **t** の組合せで表現できる．R，**t** に加え画像深度 z，カメラ行列 K が与えられれば時刻 T+ΔT の画像は時刻 T の画像を用いて表現できる．予測された時刻 T+ΔT の画像と正しい時刻 T+ΔT の画像を比較することで再現損失が計算できる．この再現損失を減らすようにモデルを訓練する．

Zhou et al. (2017) では訓練画像の K は既知とし，単画像から深度 z を予測する Depth CNN と，2 つの隣接画像の組から回転行列 R と移動ベクトル **t** を予測する Pose CNN の 2 つのニューラルネットワークを，動画像を用いて同時に訓練する．訓練時，カメラ行列 K はあらかじめ準備し，与える必要がある．

Pose CNN でカメラ行列 K の予測も行うことで動画像のみでの訓練を可能にし，かつ深度推定精度を向上させた研究として Gordon et al. (2019) がある．

また，深度推定の精度を評価する指標として，Zhou et al. (2017) と Gordon et al. (2019) では Absolute Relative Error (以下 Abs Rel Error と略記)が用いられている．Abs Rel Error は以下の手順により計算される．なお，今回扱う画像のサイズは縦 M 横 N とし，画像深度は各ピクセルの値を長さ MN のベクトルとして表現するものとする．正解画像深度を z_{gt}，Depth CNN が推定した画像深度を z_{pred} とする．

1. z_{gt} の成分のうち値が 1mm 以上 80m 未満の成分のみを取り出したベクトル z'_{gt} を作成
2. z'_{gt} に対応する z_{pred} の成分を取り出し z'_{pred} とする
3. $s=\mathrm{median}(z'_{gt})/\mathrm{median}(z'_{pred})$ と定義
4. $z''_{pred}=s\,z'_{pred}$ として予測深度と正解深度の中央値を合わせる
5. z''_{pred} の成分のうち値が 1mm 未満の要素の値を 1mm に，80m を超える要素の値を 80m に変更
6. z'_{gt} と z''_{pred} の各成分 i について $|z''_{pred,i}-z'_{gt,i}|/z'_{gt,i}$ を求め，全ピクセルについての計算値の平均をその画像の Abs Rel Error とする
7. 評価用画像すべての Abs Rel Error の平均値をその深度推定モデルの Abs Rel Error とする

手順 3 で導入される定数 s は予測深度と正解深度のスケールの差を消すための定数である．予測深度のスケールが正解深度と合っており，予測深度と正解深度の中央値が等しければ s=1 となる．本論文ではこの画像ごとに決まる定数 s の全評価画像についての平均値を定数倍誤差補正定数と呼ぶことにする．

手順 6 から，定数倍誤差補正後予測深度と正解値が近いほど Abs Rel Error は小さくなることがわかる．

表 1 に Zhou et al. (2017)と Gordon et al. (2019)の Abs Rel Error と定数倍誤差補正定数を示した．Gordon et al. (2019) の方が Abs Rel Error が小さいことから，Gordon et al. (2019) において深度推定精度が改善されていることがわかる．また，両モデルともに定数倍誤差補正定数が 1 から外れていることから，推定深度が定数倍誤差を持つことがわかる．

表 1 を作成するにあたり訓練・評価に用いられたデータセットは KITTI データセットとよばれ，道路画像深度推定の定量的評価に幅広く用いられている．KITTI データセットにはドイツの道路動画とレーザー測定深度の組が含まれている(Geiger et al., 2013).

３．実験の方法

３．１　ステレオカメラ動画の入手

2020 年 7 月 16 日，四輪自動車のフロントガラスにステレオカメラを取り付け，埼玉県狭山市と川越市にまたがる 5km 区間の往復撮影を行った．往復により計 2 本の動画を撮影した理由は，1 本(往路)を評価用動画，1 本(復路)を学習用動画としたためである．

ステレオカメラは図 1(a)のようにフロントガラスに吸盤で付着後，養生テープで固定した．カメラ加速度測定のため，図 1(b)のような座標系を持つ加速度センサ付きラズベリーパイを，ステレオカメラの下にマジックテープと養生テープを用いて固定した．

３．２　KITTI 訓練済モデル作成とパラメータ補正

今回用いるステレオカメラの出力はグレースケー

表 1 既存モデルの評価値と定数倍誤差補正定数

	Abs Rel Error	定数倍誤差補正定数
Zhou et al. (2017)	0.183	3.01
Gordon et al. (2019)	0.126	28.5

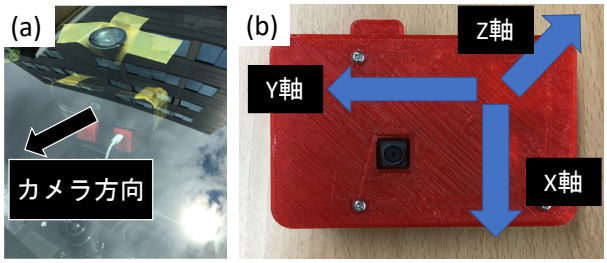

図 1 実験装置の概略
(a) ステレオカメラ設置状態とカメラ方向
(b) 加速度センサ付きラズベリーパイとその座標系

ルであるが，既存モデルはカラー画像で訓練・評価されている．そこで KITTI データセットのカラー画像をグレースケールに変換した上で訓練を行い，モデルを準備した． 作成したモデルについて，3.1 節で学習用に準備した復路動画から作成した連続画像群を学習させ，パラメータを補正した．

３．３ 相対深度推定評価

パラメータ補正前後のモデルそれぞれについて，3.1 節で評価用に準備した往路動画に含まれる全画像を用いて Abs Rel Error を計算した．深度正解データはステレオカメラ推定深度としている．なお，ステレオカメラを車両内に設置したことと，左右画像の対応点検索の失敗のため，ステレオカメラ推定深度には極端に小さな値が含まれていた．これらの値の影響を除くため，本節以降 Abs Rel Error 計算時には Abs Rel Error 計算手順中の「1mm」という値を全て「5m」と読み換えた．

３．４ 定数倍誤差補正評価

往路動画撮影中の x 方向加速度平均値を a とすると，カメラ方向と加速度センサの x 軸は平行だから，カメラの水平面からの傾き θ は重力加速度 g を用いて θ =arcsin (|a|/g)となる．カメラ設置高 h は計測車両が全高 1.490 m であることから h=1.2m と推定できるから，画像中央の絶対深度 d_{center} を d_{center}=h/sin θ と推定できる．よって，Abs Rel Error 計算手順 3 について，s=d_{center}/(推定画像深度の画像中央に対応する成分)のように s の定義を変更すれば，モデルが推定する相対深度について正解データの中央値を使わずに定数倍誤差補正を行い，絶対深度に変換できる．

上記の考え方に従い，3.2 節で作成したパラメータ補正後モデルから推定される相対深度について，Abs Rel Error 計算手順 3 の s の定義を s=d_{center}/(推定画像深度の画像中央に対応する成分) に変更し，Abs Rel Error を計算した．また相対深度推定値をそのまま絶対深度として用いた場合と比較するため，s の値を常に 1 とした Abs Rel Error 計算も実施した．

４．実験の結果

４．１ 相対深度推定評価結果

3.2 節で準備したパラメータ補正前後のモデルをそれぞれ「補正前モデル」「補正後モデル」とし，Abs Rel Error を計算した値を表 2 に示した．

具体的な深度推定例として，図 2(a)の画像をステレオカメラで深度推定した結果を図 2(b),「補正前モデル」で深度推定した結果を図 2(c),「補正後モデル」で深度推定した結果を図 2(d)に示した．色が明るいほど推定深度の値が大きく，画像内の対応する位置がカメラから遠いと推定されたことを示す．図 2(b)は本論文では正解データとして扱っているデータであり，疎データのため黒色部の値は 0 である．図 2(c)(d)は相対深度を示しているため値の絶対的大きさは意味を持たず，また暗色部も非ゼロ値を持つ．

４．２ 定数倍誤差補正評価結果

評価用動画撮影中の x 方向加速度平均値は-0.065g (g は重力加速度)であったため，カメラの水平面からの傾き θ は 3.7 度，画像中央絶対深度 d_{center} は 23m と推定された．加速度センサから推定された画像中央絶対深度を用いて定数倍補正を実施するモデルを「加速度センサ利用モデル」，s=1 に固定し定数倍誤差を補正しないモデルを「定数倍誤差未補正モデル」として，Abs Rel Error を計算した値を表 3 に示した．

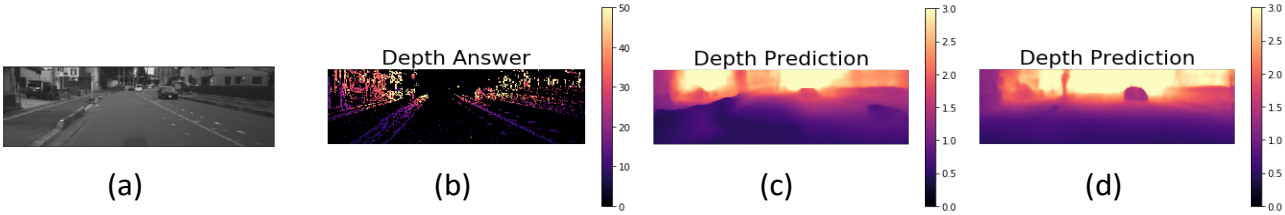

図 2 入力画像と推定結果例
(a)入力画像　(b)ステレオカメラ深度推定結果(単位:m)
(c)パラメータ補正前モデル推定結果(相対値)　(d)パラメータ補正後モデル推定結果(相対値)

表 2 復路動画による深度推定モデル補正結果

	Abs Rel Error
補正前モデル	0.280
補正後モデル	0.226

表 3 加速度センサによる定数倍誤差補正結果

	Abs Rel Error
加速度センサ利用モデル	0.403
定数倍誤差未補正モデル	0.680

５．考察

５．１　相対深度推定評価結果

表 2 の通り，KITTI データセットのみで学習させたモデルのパラメータを復路動画により補正したモデルにおいて Abs Rel Error が小さくなったことから，ドイツの動画で学習させたモデルのパラメータを日本の動画を用いて補正することの有効性が示された．

図 2(c)に比べ図 2(d)では車両がより鮮明に浮かび上がっており，定性的にも精度向上が示された．

一方，ここで計算した Abs Rel Error は，3.4 において正解値の範囲を既存研究よりも狭めたにも関わらず表 1 で示した既存研究での値より大きくなっており，改善の余地が存在することがわかる．

５．２　定数倍誤差補正評価結果

表 3 の通り，定数倍誤差補正を加速度センサで行ったモデルについて，定数倍誤差補正を行わないモデルに比べ Abs Rel Error が約 40%減少したことから，加速度センサを用いた定数倍誤差補正の有効性が示された．

一方，表 3 の「加速度センサ利用モデル」の Abs Rel Error は，共通の相対深度推定結果について正解データの中央値を用いて定数倍誤差補正を行った表 2 の「補正後モデル」の値に比べ大きく，定数倍誤差補正についても改善の余地が存在するといえる．

６．おわりに

本論文では KITTI データセット・埼玉県で撮影した車載動画・加速度センサをもとに画像深度推定モデルを構築し評価を行った．今回作成したモデルは正解データを用いずに定数倍誤差を補正可能であるため，絶対深度推定モデルとしての利用価値が見込める．一方，今回作成したモデルには相対距離推定・定数倍誤差推定の両面で改善の余地が存在する．今後はより多くの動画を用いてモデルを改良するとともに，より高精度の定数倍誤差補正方法を検討すべきであるといえる．

参考文献

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013) Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, **32**(11), 1231-1237.

Gordon, A., Li, H., Jonschkowski, R., & Angelova, A. (2019) Depth from videos in the wild: Unsupervised monocular depth learning from unknown cameras. *In Proceedings of the IEEE International Conference on Computer Vision*, 8977-8986.

Zhou, T., Brown, M., Snavely, N., & Lowe, D. G. (2017) Unsupervised learning of depth and ego-motion from video. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1851-1858.