

ドローンの自律飛行のための赤外線画像から深度画像の生成

嶋田 知泰¹ 西川 広記^{1,2} 孔 祥博^{1,a)} 富山 宏之^{1,b)}

概要: 本論文では、ドローンの夜間飛行における衝突を回避するために、赤外線カメラの画像から深度画像を生成する手法を提案する。ドローンの最高飛行速度は一般的に約 22.2m/s であり、安全に飛行するためには長距離の深度情報が必要となる。しかし、長距離測定可能な深度カメラはドローンに搭載するには重量が大きく、ドローンが安定して飛行することが困難である。本論文では、Pix2Pix という条件付き敵対生成的ネットワークの一種であるモデルを使用し、赤外線カメラから長距離測定可能な深度画像の生成を行う。モデルの学習に用いたデータセットはフライトシミュレータの 1 つである AirSim を使用し、収集したデータセットを用いて学習されたモデルによって、現実の赤外線カメラから深度画像の生成を可能としている。提案手法の有効性を PSNR と SSIM の観点から評価する。また、夜間に赤外線カメラから取得した画像を用いて深度画像を生成して視覚的な評価を行う。さらに提案手法をフライトシミュレータに活用し衝突回避に対する有効性を評価する。

キーワード: CGAN, Pix2Pix, 深度画像, ドローン, 自律飛行

1. はじめに

近年、小型ドローンは、利便性、低消費電力、低価格などの観点から一般に普及しており、その利便性を活かした様々な役割がこれまで以上に期待されている。期待されている役割の例として、インフラの点検、荷物の配送、移動式監視カメラなどがあげられる。ドローンは自動車や他の航空機などの有人の乗り物とは異なり、人が搭乗する必要がない。そのため有人航空機では多くの危険を伴うような場所での飛行であってもドローンであれば飛行が可能である。そしてそれらの利便性をさらに活かすために自律飛行の研究が盛んに行われている。ドローンの自律飛行においては、衝突回避が不可欠であり、重要な課題の一つとされている。従来の手法では、距離センサーを用いるのが一般的である。例えば、長い距離を検出できる LiDAR (Light Detection and Ranging) を搭載する手法が提案されている [1], [2]。また、距離を認識するために、深度カメラやステレオカメラなどを搭載する手法もある [3], [4], [5], [6]。しかし、このような高性能なセンサは小型のドローンに搭載すると重くなり安定な飛行が困難となる。またドローンはバッテリーで駆動するため重量が大きくなることによって

消費エネルギーが増加してしまい、長距離の飛行ができなくなるというデメリットがある。一方、低性能のセンサでは遠距離を高精度に見ることができず、消費エネルギーを低減することができるが、物体との衝突のリスクが高まる。そのため軽量の単眼カメラを用いてそれらの高性能センサの代替となるような手法が提案されている。例として、単眼カメラからドローン周辺の物体を検出・認識することや深度カメラのように動作するように深度推定手法などがあげられる [7], [8]。しかしながら、これらの手法は日中の飛行を前提としているものが多く、夜間に単眼カメラから特徴量の抽出が難しいため正常に動作しない。そこで本論文では赤外線カメラを用いる長距離深度画像の生成手法を提案する。本論文の貢献は以下の 3 点である。

- 夜間におけるドローン自律飛行のために、赤外線画像から長距離の深度画像を生成する手法は我々が知る限り、初めての論文である。
- 長距離測定可能な深度画像をリアルタイムで推定できることを検証し、フライトシミュレータでドローンが障害物を避けて飛行できることを実証している。
- 産業上の実用性の観点から、提案した手法を実在する赤外線カメラに適用し、精度の高い深度画像が生成できることを確認している。

本論文の構成は以下のとおりである。2 章では本研究の関連研究について述べる。3 章では赤外線画像から深度画

¹ 立命館大学大学院理工学研究科
Ritsumeikan University, Kusatsu, Shiga, Japan

² 日本学術振興会特別研究員 DC1
JSPS Research Fellows

a) kong@fc.ristumei.ac.jp

b) ht@fc.ristumei.ac.jp

像の生成手法について述べる。4章では実験結果を示し、5章で本論文のまとめと今後の課題について述べる。

2. 関連研究

ドローンの自律飛行に関する研究は、数十年前から数多く行われている。中でも、特に物体との衝突を回避するための安全飛行に焦点が当てられている。論文 [1], [2], [9] では、著者らは、ドローンの障害物回避に LiDAR を使用している。しかし、大量のセンサーや高性能なセンサーを搭載すると、ドローンの重量が増加し、エネルギー消費量が増加する。また、ドローンの飛行にはバッテリー容量の制限があるため、多数のセンサーを搭載したまま長距離飛行することは困難である。

この問題を解決するために、論文 [3], [4], [6] のアプローチでは、軽量・小型の深度カメラやステレオカメラを用いた衝突回避手法を提案している。深度画像を用いて最適な飛行方向を決定することで、ドローンが飛行中に障害物を回避することができる。しかしながら一般的にドローンに搭載可能な小型深度カメラでは高々 20 メートルしか測定できない。そのため障害物を避けながら安全に飛行するには低速飛行が強いられる。これらのことから安定して長距離を飛行するために、我々は軽量の単眼カメラを用いた長距離深度推定の手法を提案する。

深度推定の関連研究を示す。単眼カメラを用いた長距離深度推定を行う研究は論文 [7] では、著者らは、サポートベクターマシン (SVM) を用いた深度推定の手法を提案している。これらのシステムは、ドローンからの画像を細かなパッチに分割し、各パッチを手で作成した特徴量のセットを用いて表現し、事前に学習した SVM 分類器を用いて各パッチの深さを推定するものである。しかし、これらのシステムは、学習データが少ないため、精度が低い。論文 [10], [11], [12], [13] では、これらの手法は Convolutional Neural Networks (CNN) に基づいている。論文 [7] では、CNN を用いた手法は SVM を用いた手法よりも精度が高いが、処理時間が長くなってしまいうため、衝突のない安全な飛行を実現するにはまだ十分ではない。論文 [8] と論文 [14] では、単眼カメラを用いた深度推定の前にセグメンテーション画像を前処理する手法を提案している。この手法を用いることで、深度推定の精度を向上させることができる。その一方で、処理負荷が大きくなるため、性能面でリアルタイム処理には適していない。論文 [15] では、著者らは深度推定の前に ORB-SLAM を用いてスパースな深度画像を生成し、その深度画像と単眼カメラの画像を合成して深度を推定する手法を提案している。しかし、夜間飛行では、ORB-SLAM や深度推定のための特徴点の取得が困難である。

そこで夜間においても深度推定を行うことができる赤外線カメラを用いた手法がある。論文 [16] では赤外線画像と

深度画像のデータセットがないため、自己教師付き深層学習フレームワークを用いた深度推定手法を自動車の自律走行向けに著者らは提案した。単眼赤外線カメラのビデオフレーム間の再現関係を用いて損失関数を構築することで自己教師あり学習を可能としている。著者らはこの手法において 15 メートル以内の深度を正確に推定できると主張している。ドローンの場合、1 秒間に 10~20 メートル程度の速度で飛行するため、この手法で推定される距離は物体との衝突を避けるには短い。また、ドローンは一般車両に比べてさらに小さく、大型のコンピュータを搭載できないため、自動車に比べて処理能力が相対的に低い。本論文では、ドローンの夜間飛行のための新しい深度推定画像生成手法を提案する。提案手法は、単眼の赤外線カメラを用いて長距離を推定することができる。推定モデルには、条件付き敵対的生成ネットワーク (CGAN) [17] を用い、学習にはドローンの仮想飛行環境として知られる AirSim [18] のデータを使用する。

3. Pix2Pix による深度画像生成手法

3.1 深度画像生成手法の概要

赤外線から深度画像の生成には CGAN [17] の一種である Pix2Pix を用いる [19]。図 1 に Pix2Pix で使用される CGAN の構成を示す。図 1 のように CGAN は生成器である Generator と、判別器である Discriminator の 2 つのネットワークから成る。Generator は Discriminator 生成された画像であると判別されないように学習し、Discriminator は Generator によって生成されたデータを学習データであると誤認しないように学習する。Generator には局所的な特徴量の抽出と位置情報の復元が可能となる U-Net [20] を用いる。

CGAN の目的関数は (1) 式のようになる [19]。

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (1)$$

$D(x, y)$ は学習データを学習データと判断する確率、 $D(x, G(x, z))$ は生成された画像を学習データと判断する確率である。 $G(x, z)$ は Generator による生成画像である。 z はノイズベクトルである。 z は必須ではないが、 z を入力せずに学習を行うと、汎化性能が悪くなる。そのため、ノイズベクトルを学習の際に入力し学習データ以外での入力であってもある程度の精度を持った出力を行うことができる。Discriminator はこの目的を最大化しようとし、Generator は最小化しようとする。そのため Generator は Discriminator を騙すだけでなく、正解に近づく画像を生成する必要がある。そのためには、CGAN の目的関数に (2) 式のような L1 ノルムを加えることが有効である [19]。

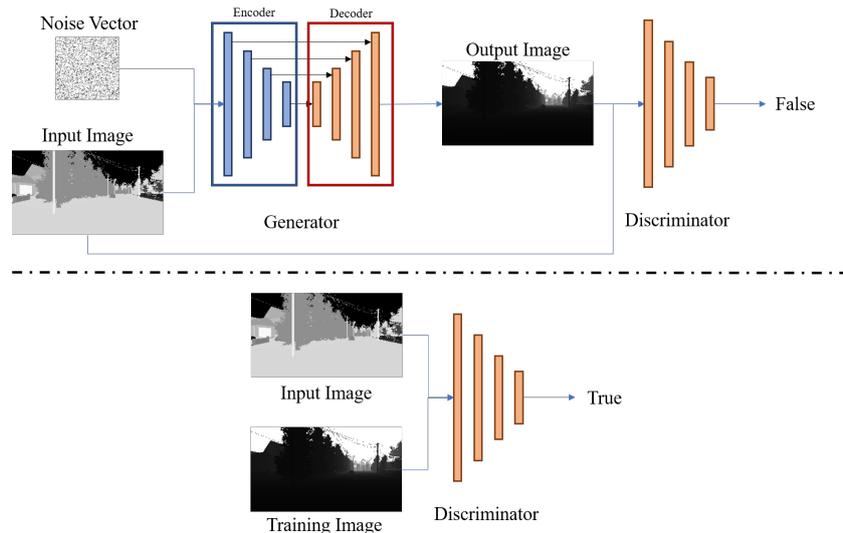


図 1 CGAN の概要

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x,z)\|_1] \quad (2)$$

L1 ノルムベースの画像生成では、画像全体を捉えることは可能であるが、細部の不鮮明さが問題となる。一方で、CGAN ベースの画像生成では、画像全体を捉えることはできないが、細部を捉えることができる。この2つの手法を組み合わせることで、精度の高い画像を生成することが可能となる。そこで、Pix2Pix の目的は (3) 式ようになる。なお、 λ は L1 ノルムの重みである。このパラメータは学習時に設定することができる。

$$G^* = \arg \min_G \max_D \mathcal{L}_{CGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (3)$$

4. 実験

4.1 実験準備

提案した手法の有効性を評価するために、実験を行った。まず、提案手法が仮想環境で深度画像を生成できることを検証する。次に、産業上の観点から実用性を評価するために、実世界の赤外線画像に提案手法を適用し、生成された深度画像を検証する。また、生成された深度画像を用いて提案手法を衝突回避のための飛行に活用し、飛行の衝突率を評価する。

実験では、AirSim[18] から取得した赤外線画像と深度画像を用いてモデルの学習を行う。AirSim は、仮想マップ上で赤外線画像と深度画像を同時に撮影できる飛行環境として、これまで以上に注目を集めている。また、AirSim 上のドローンに、対象物の位置を示す地図の情報を静的に与えることで、100メートル以上の画像を仮想的に撮影することができる。例えば、AirSim で得られた深度画像は、図2に示すように、長い距離を描写することができる。学習には、「City」、「Coastline」、「Neighborhood」、「Soccer

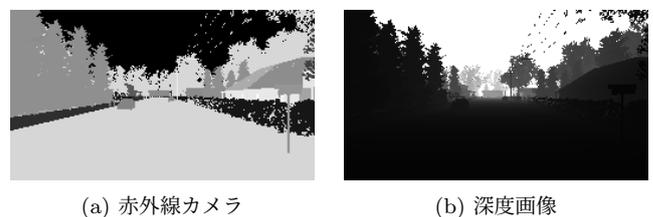


図 2 AirSim で得られるデータセットの例

表 1 トレーニングで用いた計算機的环境

Epochs	400
Pairs	1600 (400 pairs for each map)
Batch Size	1
L1 weight	100
OS	Ubuntu 20.04 LTS
RAM	32GB
CPU	Intel Core i9-10900K
GPU	NVIDIA GeForce RTX 3090

Field」と呼ばれる4つのマップ上で、赤外線画像と深度画像を集めデータセットとした。

表1に示すように、次のように表される環境でモデルを学習を行う。エポック数は400とする。赤外線画像と深度画像のペアの数は1,600としている。バッチサイズは1、L1 ノルムの重みは100に設定している。4つのマップそれぞれについて、1,600組の赤外線画像と深度画像を集めてモデルを学習し、2,000枚の赤外線画像を用いてモデルをテストする。

4.2 精度と処理時間の評価

本提案が衝突回避に適用できるほど鮮明な深度画像を生成できることを検証するために AirSim のデータを用いて類似性の検証を行う。図3に AirSim の各マップで得られる赤外線画像を入力した際の出力の例を示す。図3(a)は

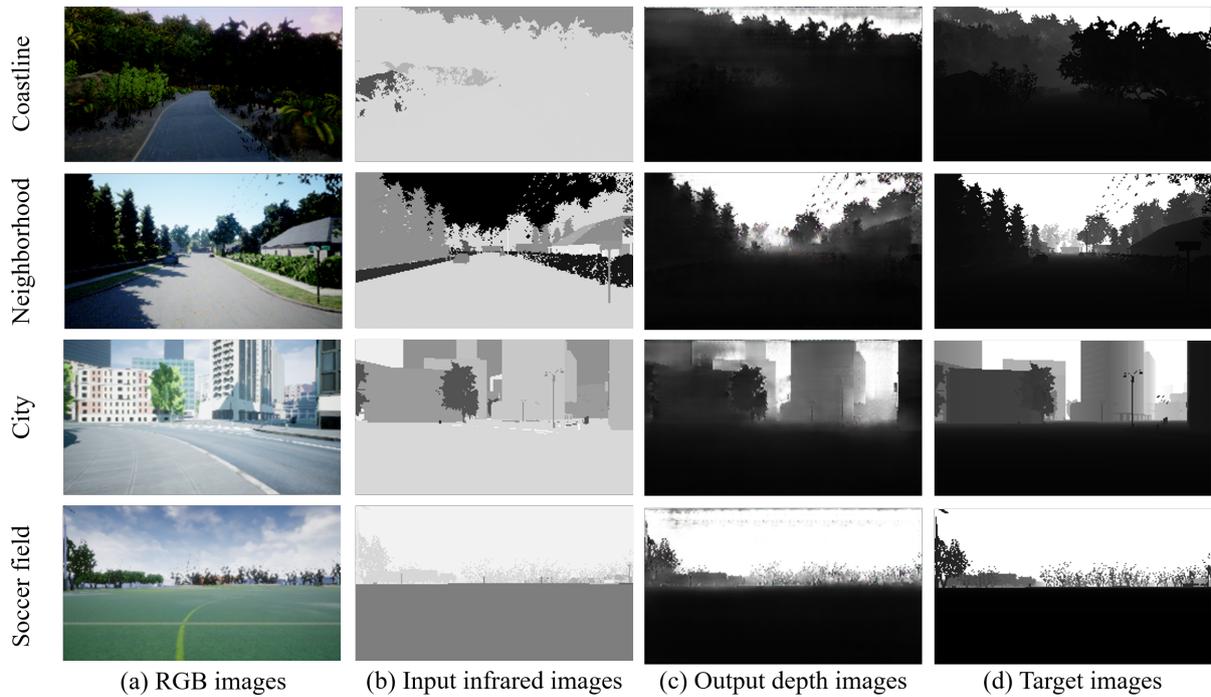


図 3 AirSim で得たデータセットを用いた際の出力

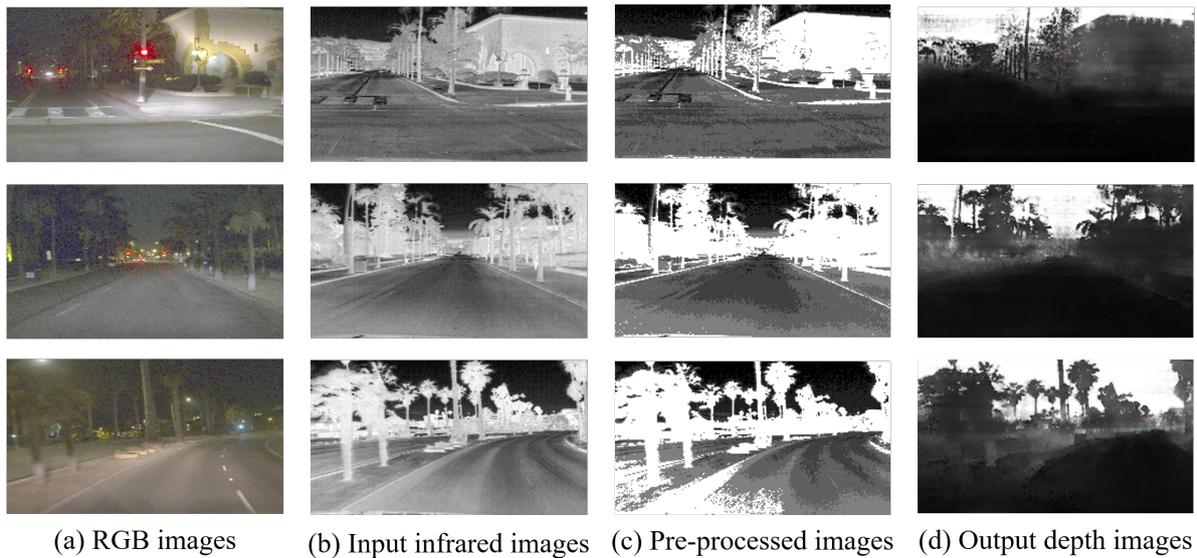


図 4 FLIR のデータセット [21] を用いた際の出力

そのマップの RGB 画像である。図 3 (b) のような赤外線画像をモデルに入力した際、図 3 (c) のような深度画像が得られる。図 3 (d) はターゲットの深度画像である。

生成された画像の品質とターゲット画像との類似性を評価するために、出力画像とターゲット画像の PSNR と SSIM の関係を使用する。PSNR が高いほど画質が高いとされ、30dB 以上であれば良いとされる。一方、SSIM は、2 つの画像の類似性を測定するために使用される、よく知られた品質指標である。SSIM が 1 に近ければ、画像は高い類似性であると評価できる。なお、我々の目的は生成された深度画像を用いて衝突を回避することであるため、画像

表 2 生成画像と教師画像の PSNR と SSIM

Maps	PSNR(dB)	SSIM
City	24.12	0.882
Coastline	25.60	0.901
Neighborhood	19.67	0.707
Soccer Field	22.88	0.834

の品質には必ずしも最重要とは限らない点に注意したい。
表 2 は、各マップの PSNR と SSIM の平均値を示している。全マップの平均の値は PSNR は 23.06, SSIM は 0.83 という結果である。PSNR の目標値の 30 には及んでいない。また、Neighborhood の PSNR が 19.67, SSIM 0.707

表 3 画像生成にかかる処理時間

Device	Processing Time(s)
NVIDIA RTX 2070 SUPER	0.031
Intel Core i7 9700K	0.183
Jetson Xavier NX	0.201

表 4 実験を行った計算機の実環境

OS	Windows 10 pro
RAM	32GB 2666MHz
CPU	Intel Core i7-9700K 3.60GHz
GPU	NVIDIA GeForce RTX 2070 SUPER 8GB
Version of Airsim	1.4.1

であり他のマップよりも悪いことがわかる。Neighborhoodでは、木や家など温度の近いものが、赤外線画像の中で重なって見えることがあり我々のモデルでは、それらを別のオブジェクトとして区別できないことがある。

また、提案手法は、ドローンが飛行中に物体との衝突を回避するために利用される。そのため、リアルタイム性という観点から、我々のモデルの処理時間を評価している。NVIDIA RTX 2070 SUPER, Intel Core i7 9700K, Jetson Xavier NX (GPU) を用いて評価を行う。表 3 は、画像ごとの処理時間を示したものである。最も遅い処理時間を示しているのは Jetson で、0.201s を表している。つまり、Jetson では 1 秒間に 5 枚のフレームを処理することができる。一方、NVIDIA RTX 2070 SUPER では 1 画像あたり 0.031s という結果になった。衝突回避のための結果の検証は、我々のモデルが生成された深度画像の中でどれだけ長く距離を推定できるかにかかっている。

最後に FLIR[21] から提供されている実際の赤外線画像を我々のモデルに入力し深度画像を生成する。AirSim の赤外線画像はオブジェクトごとに温度を割り当てている。そのため実際の赤外線画像と差異がある。そこで実際の赤外線画像を AirSim で得られる赤外線画像に近づけるために前処理を行う。前処理では 5 値化を行う。赤外線画像の輝度値を 5 分割し 5 つの閾値を決定する。その後それぞれの閾値において値を正規化することによって AirSim の赤外線画像に近づける。図 4 に実際の赤外線画像を我々のモデルに入力したときの結果を示す。図 4 (a) は RGB 画像であり、図 4 (b) は赤外線画像である [21]。図 4 (c) は、前述した前処理を施した後の赤外線画像であり、それを入力した時のモデルの出力画像が図 4 (d) である。この結果から提案モデルは、精度の高い長距離の深度画像を生成できることがわかる。RGB 画像と比較して、深度画像では物体の位置が明確に分かる。しかしながら、データセットには距離情報が含まれていないため、PSNR と SSIM をそれぞれ評価することはできない。しかし、視覚的におよその遠近がわかる深度画像の生成に成功している。

表 5 各マップにおける衝突率

Maps	Collision Rate(%)
City	21.45
Coastline	0.00
Neighborhood	0.00
Soccer Field	0.00

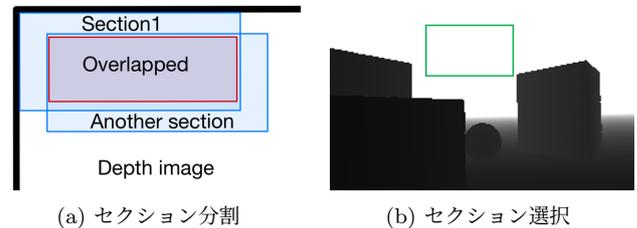


図 5 セクションの分割方法と選択 [6]

4.3 衝突回避の評価

前節まで、提案手法の精度と処理時間を評価してきた。ここでは、AirSim を用いてドローンの飛行シミュレーションを行い、提案手法が物体との衝突を避けて飛行できることを実証する。自律型ドローンの安全な飛行を実現するためには、自ら経路を計画すること、すなわちドローンが物体との衝突を回避できるように方向を選択する必要がある。実験では、論文 [6] で提案されている経路計画手法を使用している。論文 [6] で紹介されているのは、深度画像を分割し方向決定する手法である。この手法では、図 5 のように、深度画像を 289 個のオーバーラップしたセクション (17 行×17 列) に分割する。オーバーラップしたセクションに分割することで、広い範囲を数多く参照することができるため、障害物を避けて安全に通過するために最適なセクション選択をすることができる。選択するセクションは画素値の合計が最も高いセクションである。実験では表 4 に示すように、主記憶 32GB, CPU に Intel Core i7-9700K(8 コア, 8 スレッド, 4.90GHz), GPU に NVIDIA GeForce RTX2070 SUPER(8GB), Windows 10 Pro を搭載した計算機を使用している。実際のドローンの飛行において RTX2070 SUPER のような高性能の GPU を搭載することができないため、実験では Pix2Pix の処理は CPU 上で実行する。飛行シミュレーションは、「City」、「Coastline」、「Neighborhood」、「Soccer Field」の 4 つのマップで行い、各マップ 400 回ずつ異なる座標を目的地として実施する。その飛行中に一度でも障害物に衝突した場合は、衝突したと判断する。

表 5 は、AirSim での飛行シミュレーションによる各マップの衝突率を示したものである。この結果から、「Coastline」、「Neighborhood」、「Soccer Field」では、提案手法はどのオブジェクトにも衝突しないことがわかる。一方、「City」では 21.4% の衝突率となっている。「City」での出力画像の精度は表 3 より「Neighborhood」よりも PSNR,

SSIMが高いにも関わらず、衝突率は高くなっている。また、「City」は他の3つのマップと比べ高層ビルが多くあるマップとなっている。これらのことから、用いた経路計画手法が高層ビルのあるような環境に適していないのだと考えられる。

5. おわりに

本論文では、夜間飛行のための赤外線画像からの深度推定法を提案する。提案手法は、CGANの1つであるPix2Pixをベースにしており、トレーニングデータセットはAirSimの仮想フライトシミュレータから収集した。提案手法により生成された深度画像の品質と類似性、リアルタイム性を実現するための処理時間、衝突率をAirSimでのフライトシミュレーションにより評価した。衝突率の結果から、提案手法は衝突回避のために十分な精度を高速で実現できることが示された。

今後は、複雑なマップ上の障害物を避けて飛行するための経路計画手法を実現する。また、精度と遅延を改善するために、新しいネットワークを設計する予定である。

謝辞 本研究の一部は科研費20J21208, 20K23333の支援による。

参考文献

- [1] A. Moffatt, E. Platt, B. Mondragon, A. Kwok, D. Uryeu, and S. Bhandari, "Obstacle Detection and Avoidance System for Small UAVs Using a LiDAR," in *International Conference on Unmanned Aircraft Systems*, 2020.
- [2] Y. Hou, Z. Zhang, C. Wang, S. Cheng, and D. Ye, "Research on Vehicle Identification Method and Vehicle Speed Measurement Method Based on Multi-rotor UAV Equipped with LiDAR," in *International Conference on Advanced Electronic Materials, Computers and Software Engineering*, 2020.
- [3] J. Borenstein and Y. Koren, "The Vector Field Histogram-Fast Obstacle Avoidance for Mobile Robots," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 278–288, 1991.
- [4] C. Ma, Y. Zhou, and Z. Li, "A New Simulation Environment Based on AirSim, ROS, and PX4 for Quadcopter Aircrafts," in *International Conference on Control, Automation and Robotics*, 2020.
- [5] D. Ma, A. Tran, N. Ketii, R. Yanagi, P. Knight, K. Joglekar, N. Tudor, B. Cresta, and S. Bhandari, "Flight Test Validation of Collision Avoidance System for a Multicopter using Stereoscopic Vision," in *International Conference on Unmanned Aircraft Systems*, 2019.
- [6] E. Perez, A. Winger, A. Tran, C. Garcia-Paredes, N. Run, N. Ketii, S. Bhandari, and A. Raheja, "Autonomous Collision Avoidance System for a Multicopter using Stereoscopic Vision," in *International Conference on Unmanned Aircraft Systems*, 2018.
- [7] K. Bipin, V. Duggal, and K. Madhava Krishna, "Autonomous Navigation of Generic Monocular Quadcopter in Natural Environment," in *International Conference on Robotics and Automation*, 2015.
- [8] A. Atapour-Abarghouei and T. P. Breckon, "Monocular Segment-Wise Depth: Monocular Depth Estimation Based on a Semantic Segmentation Prior," in *IEEE International Conference on Image Processing*, 2019.
- [9] J. Redding, J. Amin, J. Boskovic, Y. Kang, K. Hedrick, J. Howlett, and S. Poll, "A Real-Time Obstacle Detection and Reactive Path Planning System for Autonomous Small-Scale Helicopters," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2007.
- [10] S. Zhang, N. Li, C. Qiu, Z. Yu, H. Zheng, and B. Zheng, "Depth Map Prediction from A Single Image with Generative Adversarial Nets," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 14 357–14 374, 2020.
- [11] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, pp. 2024–2039, 2016.
- [12] M. Mancini, G. Costante, P. Valigi, and T. A. Ciarfuglia, "J-MOD2: Joint Monocular Obstacle Detection and Depth Estimation," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1490–1497, 2018.
- [13] K. Hatch, J. Mern, and M. Kochenderfer, "Obstacle Avoidance Using a Monocular Camera," in *AIAA Scitech Forum*, 2021.
- [14] Q. Hou and C. Jung, "Occlusion Robust Light Field Depth Estimation Using Segmentation Guided Bilateral Filtering," in *IEEE International Symposium on Multimedia*, 2017.
- [15] X. Yang, J. Chen, Y. Dang, H. Luo, Y. Tang, C. Liao, P. Chen, and K.-T. Cheng, "Fast Depth Prediction and Obstacle Avoidance on a Monocular Drone Using Probabilistic Convolutional Neural Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 156–167, 2021.
- [16] X. Li, M. Ding, D. Wei, X. Wu, and Y. Cao, "Estimate Depth Information from Monocular Infrared Images Based on Deep Learning," in *IEEE International Conference on Progress in Informatics and Computing*, 2020.
- [17] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [18] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles," in *Field and Service Robotics*, 2017. [Online]. Available: <https://arxiv.org/abs/1705.05065>, [Accessed On Jun., 2021].
- [19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional Networks for Biomedical Image Segmentation," in *IEEE International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.
- [21] FLIR, "FREE FLIR Thermal Dataset for Algorithm Training," [Online]. Available: <https://www.flir.jp/oem/adas/adas-dataset-form/>, [Accessed On Sep., 2021].