

[研究助成 (A)]

災害現場での救援活動を支援する超小型 3 次元再構成センサの開発

2241015

研究代表者
(助成金受領者)

法政大学

教 授

小 西 克 巳

共同研究者

東京工科大学

講 師

佐々木 亮 平

[研究の目的]

災害発生時に災害現場の被災状況を早急に把握することは、救援活動における重要課題である。特に人が立ち入ることが困難な屋内等の災害現場では、二次災害を避けるためにも必要不可欠である。近年では、カメラ映像や深度センサ等のセンサ情報から周囲の 3 次元地図を作成する技術が発展してきており、災害現場にドローンやロボットを投入することで 3 次元地図を作成し、災害対応計画を作成することが可能となる。しかしながら、詳細な 3 次元地図を作成するには、カメラを搭載したロボットやドローンが必要となり、コストが高く準備に時間も要するため、実際の災害現場での導入の障壁となっている。そこで本研究では、災害現場に人の手で投げ込むだけで、高精度な 3 次元地図を作成する安価で使い捨て可能な手のひらサイズ以下のボール型 3 次元再構成センサの開発を目指す。災害現場において、人による救援活動、および、人と自律ロボットが協働した救援活動を実現するには、災害現場の詳細な情報を知ることが必要不可欠である。超小型 3 次元再構成センサの実現により、容易に 3 次元地図が作成され、迅速な災害対応計画の立案が可能となる。

本研究で開発を目指す超小型 3 次元再構成センサは、手のひらサイズ以下の小型のボール型カメラで、災害現場に複数個を投げ入れて動画を取得し、その動画像群から超解像度 3 次元地図を生成するものである。本研究の目標は、

ボール型カメラから得られる動画像群から 3 次元地図を作成するソフトウェアを開発するための数理基盤の確立である。カメラ映像のみから 3 次元地図を作成する技術は SLAM と呼ばれ、多くの手法が開発されている。しかし、これらの手法は自動車やロボット等に設置されたカメラで撮影された滑らかに動く映像を想定されている。本研究の目的は、時間方向にも空間方向にも疎にサンプリングされたデータから 3 次元地図を作成する手法の確立である。

[研究の内容、成果]

提案手法の概要

本研究では、災害現場で疎にサンプリングされたデータから 3 次元地図を作成する手法の確立を目指すため、赤外線画像を用いて照明が少ない条件下でも安定した深度補完を行う手法を提案する。RGB ガイド付き手法に着目し、ガイドに RGB 画像ではなく赤外線画像を入力に用いて深度補完手法を構築した。提案手法の全体アーキテクチャ以下に示す。

ネットワーク構造は PENet を参考にし、一

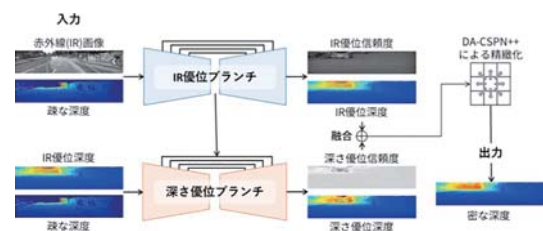


図1 提案手法のアーキテクチャ

方のブランチを赤外線情報に依存した深度を予測する IR 優位ブランチとすることで赤外線画像をガイドに用いた補完を実現した。具体的には、入力チャンネルを RGB+深度の 4 チャンネルから赤外線+深度の 2 チャンネルに減らし、計算の効率化を図った。また、実際のハードウェアに組み込まれることを想定して、実行時間の短縮を目指している。学習データには、深度のグラントゥルースおよび赤外線画像を使用した。実験では、様々な照明条件下で取得されたデータセットを用いて、既存の RGB ガイド付き手法との精度および実行時間の比較を行い、本手法の優位性を示した。

2 分岐バックボーン

2 分岐バックボーンは、それぞれのブランチから得られる IR（赤外線）優位の情報と深さ優位の情報を徹底的に利用し、2 つのモダリティの融合を効果的に行う方法である。2 つのブランチでは、同様なエンコーダ・デコーダネットワークを構築し、IR 優位の深度予測と深さ優位の深度予測を行う。IR 優位のブランチは、まず赤外線画像から高密度の深度マップを予測することを目的とする。深度予測の有効性を高めるために、疎な深度マップも入力される、このブランチは、スキップ接続を持つエンコーダ・デコーダ構造のネットワークである。エンコーダは 1 つの畳み込み層と 10 個の基本残差ブロック、デコーダは 5 つの逆畳み込み層と 1 つの畳み込み層を持つ。すべての畳み込みの後に BatchNorm と ReLU を行う。このブランチには赤外線画像と疎な深度マップの両方が入力されるが、このブランチは深度予測のために IR 優位の特徴を抽出し、赤外線画像の構造情報を利用してオブジェクト境界付近の深度を学習できるようにする。深さ優位のブランチは、まず疎な深度マップをアップサンプリングすることで、密な深度マップを予測することを目的とする。このブランチでは、同様のエンコーダ・デコーダネットワークが構築される。さら

に、このブランチに IR 優位ブランチの特徴量を融合させる。具体的には、IR 優位ブランチのデコーダ特徴を深さ優位ブランチの対応するエンコーダ特徴と連結する。さらに、IR 優位ブランチで得られた深度予測結果もこのブランチに入力される。このようにして赤外線と深度の特徴が複数の段階で融合される。最後に 2 つのブランチで予測された密な深度画像を融合する。

幾何学的畳み込み

提案手法では、先行研究と同様に 3D 幾何学的情報を符号化するために幾何学的畳み込み層を採用している。この手法によって、カメラ座標における 3 次元位置をレイヤーの入力に連結することで、従来の畳み込みレイヤーを単純に補強することができる。位置マップ (X, Y, Z) は、元の疎な深度マップから導出される。この手法により、3 次元幾何学的情報を 2 つのブランチ特徴により効率的に符号化することができる。

CSPN++ の拡張と高速化

深層ニューラルネットワークによって生成された深度画像は、入力深度値があるピクセルにおいて入力値を保持しない可能性がある。SPN は有効なピクセルにおける深度値を回復するために隣接画素間の親和性を学習することで、推定された深度画像の精緻化を行う。本手法では、SPN に基づく畳み込み精緻化手法である CSPN++ を採用し、バックボーンによって予測された深度マップを改良する。さらに、CSPN++ をベースにより効率を上げるための伝播近傍の拡大と伝播の並列実装を行い、伝搬手順を大幅に高速化した DA-CSPN++ を採用する。

実験

・環境

実装は Pytorch で行い、NVIDIA GeForce

RTX 3090 で学習を行った。オプティマイザには Adam を用いた。PENet, 提案手法は重み減衰は 10^{-6} とする。バッチサイズ 6, 初期学習率 0.001, 10 エポックで学習する。画像を 576×288 にランダムクロップする。モデルの学習には約 20 時間を要する。

・学習方法

ネットワークのトレーニングには、データセットから取得したガイド画像と深度画像を使用する。ガイド画像にはカラー画像, モノクロ画像, 赤外線画像を用いる。疎らな深度画像は, 従来の研究と同様にグランドトゥルースの有効画素から 70% の画素を欠損させて生成する。疎らな深度画像には 1.25%, グランドトゥルースには 4.17% の有効深度値が含まれている。

・データセット

提案する方法でネットワークを学習するためには, 同一シーンで撮影された RGB, 深度, 赤外線データを含むデータセットを必要とする。提案手法が Visual SLAM や行動認識などのタスクに応用されることを想定し, 本研究では ViViD++ データセットを使用する。

・ViViD++ データセット

ViViD++ データセット [6] は, ロバストな Visual SLAM を開発するために作成され, 手持ち式または自動車のセンサーから, 異なる条件で繰り返し記録されたデータセットである。センサーは RGB, 赤外線, 深度, イベント (輝度の瞬間的な時間変化) を測定することで, 可視光強度に左右されないデータを得ることができる。データセットは, 異なる照明条件をもつ屋内シーケンスと屋外シーケンスを持つ。

・データセット生成

ViViD++ データセットは ROSbag 形式で提供されているため, 前処理を行い点群データから学習データセットとテストデータセットを生成する。まず LiDAR による点群データをカメラ行列, 歪み行列を用いて変換し, カラー画像または赤外線画像とマッチする深度画像を生成する。その後リサイズとクロップを行う。

具体的には $1/2$ にダウンサンプリングした後, 640×320 にクロップする。本研究で使用する屋外データセットには 8 シーン 55987 フレームが含まれ, そのうち 44782 フレームが学習データに, 11205 フレームがテストデータに用いられる。また, 使用する全 8 シーンのうち昼に撮影された 4 シーケンスを昼間シーン, 夕方と夜間に撮影された 4 シーンを夜間シーンとし, 各シーンのみでも実験を行う。

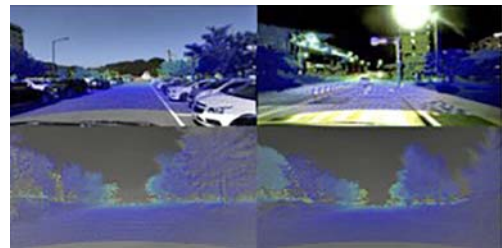


図2 ViViD++ データセットから生成した深度画像

・評価指標

性能の評価は, PENet をはじめとした多くの論文で採用されている KITTI 深度補完ベンチマークと同じく, 平均二乗誤差 (RMSE), 平均絶対誤差 (MAE), 平均相対誤差 (REL) の 3 つの指標を採用する。

実験結果と考察

実験における定性的結果を図 3 と図 4 に示す。左から順に, ガイド画像, 疎な深度画像, 補完した深度画像, グランドトゥルースである。これらの図から分かるように, どちらの手法も極めて少ない深度情報から, ガイド画像に従って深度補完が行えていることがわかる。しかし, どちらの結果も空などの深度が欠落している領域にノイズが発生している。この原因は, 使用したデータセットのデータ数や深度サンプル数が少なかったことだと考えられる。深度補完分野の研究に広く用いられている。KITTI 深度補完データセットに含まれる有効画素の割合は約 16% であるのに対し, 本研究で作成した深度画像は 4.17% である。また, KITTI データセットのデータ数は約 9.5 万であるのに対し,

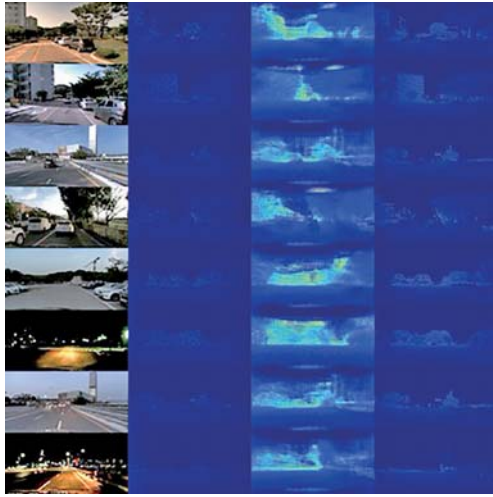


図3 定性的結果 (PENet, RGB)

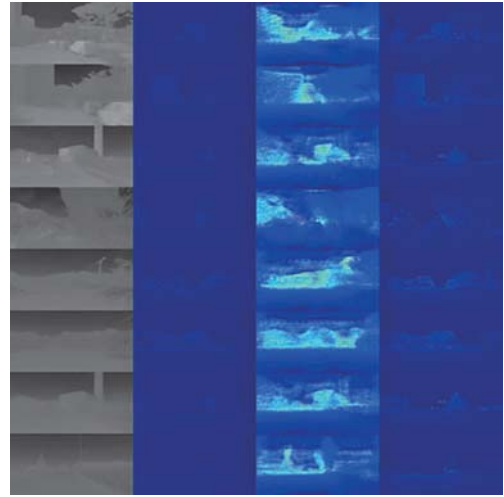
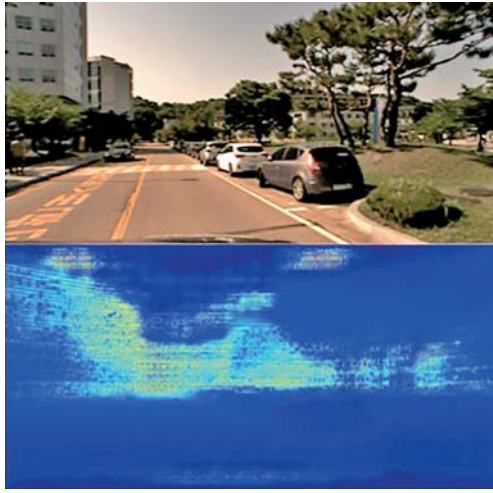
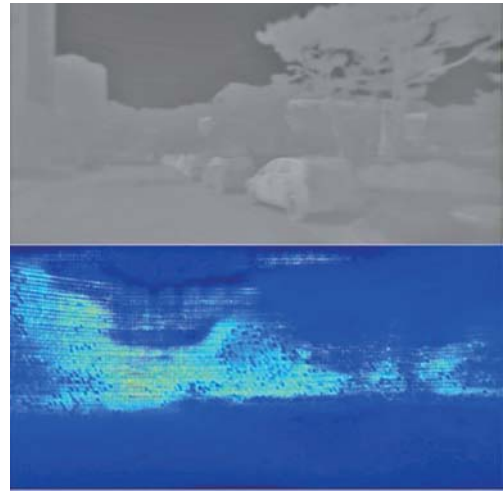


図4 定性的結果 (提案手法, Thermal)



作成したデータセットのデータ数は約5万である。したがって、ノイズの少ない補完を行うためには、より高密度なデータセットを用いる必要があると考察する。

実験における定量的結果を表1に示す。RMSEにおいては、どのシーンにおいても赤外線を入力した提案手法が最もよい結果を示している。カラー画像を入力したPENetと比較すると昼間のデータでは7.18%，夜間のデータでは8.75%，全体では7.93%RMSEが改善した。しかし、MAEおよびRELでは、どのシーンにおいても赤外線を入力したPENetが最もよい結果を示している。MAEやRELと比較して、RMSEは外れ値に左右されやすい性質がある。このことから、提案手法における結果はPENetと比較するとノイズが多く発生してい

表1 定量的結果

シーン	手法	入力	RMSE ↓	MAE ↓	REL ↓	sec/itr
全て	PENet	RGB	924.663	251.880	0.043	0.020
		GRAY	930.393	251.220	0.042	0.020
		Thermal	853.985	230.731	0.038	0.024
	提案手法	GRAY	915.527	246.907	0.044	0.023
		Thermal	851.344	<u>233.865</u>	<u>0.040</u>	<u>0.022</u>
昼間	PENet	RGB	888.982	245.666	0.045	<u>0.020</u>
		GRAY	891.149	244.817	0.044	0.019
		Thermal	<u>827.432</u>	226.792	0.040	<u>0.020</u>
	提案手法	GRAY	874.117	240.801	0.046	<u>0.020</u>
		Thermal	825.171	<u>230.198</u>	<u>0.042</u>	0.019
夜間	PENet	RGB	967.145	259.279	0.041	0.020
		GRAY	977.116	258.820	0.040	0.020
		Thermal	885.599	235.421	0.035	0.020
	提案手法	GRAY	964.830	254.175	0.042	0.020
		Thermal	882.505	<u>238.229</u>	<u>0.038</u>	0.020

と考えられる。しかしながら、どの誤差指標でも赤外線を入力に用いたモデルが最もよい結果であり、ガイド付き深度補完における赤外線情報の有用性を示しているといえる。

むすび

本研究では時間方向にも空間方向にも疎にサンプリングされたデータから3次元地図を作成する手法の導出を行なった。災害現場で疎にサンプリングされたデータから3次元地図を作成する手法の確立を目指すため、赤外線画像を用いて照明が少ない条件下でも安定した深度補完を行う手法を提案した。

赤外線カメラからの画像と、それに対応した低密度の深度データを入力し、高密度の深度データを推定する深層畳み込みニューラルネットワークの実装を行った。提案手法は、2つのオートエンコーダから深度を推定するPENetに照明条件にロバストな赤外線情報を入力することで、撮影時のシーンの照明に左右されない深度補完が可能となる。実験結果から、赤外線

画像を用いた深度補完手法は、カラー画像を用いたものよりも高い精度を達成した。特に、夜間のデータにおける補完では、昼間のデータの結果と比較して精度の改善率が高くなる結果となった。今後の課題としては、ノイズの発生を抑え、赤外線情報を最大限に利用するために、学習率などのパラメータの修正や多様で広範なデータセットを使用する必要がある。加えて、他のネットワークを適用することや、赤外線以外のモダリティを利用することも検討の余地がある。本研究により、照明条件に影響されずに深度補完を行うことが可能となる。この手法をハードウェアに組み込むことにより、暗所での深度測定がより正確になり、ロボットやドローンの物体検知システムなどでの活用が期待される。

[成果の発表、論文など]

清水大輔, 小西克巳, “赤外線画像を用いた照明条件にとらわれない深度補完”, 電子情報通信学会総合大会, 2025年3月