

Title	Vision transformerを用いた高品質かつ高速なnon-line-of-sightシーン再構成手法
Sub Title	
Author	菅野, 真央(Isogawa, Mariko) 五十川, 麻理子
Publisher	慶應義塾大学AI・高度プログラミングコンソーシアム
Publication year	2023
Jtitle	AICカンファレンス予稿集 (2023.) ,p.37- 38
JaLC DOI	
Abstract	Non-Line-of-Sight(NLOS)イメージングとは、コーナー越しに遮蔽された物体の形状を復元する技術である。光を壁に照射し、その反射光が被遮蔽領域にある物体を照らす。その物体から反射した光が壁、そしてセンサへと戻ってくる。観測される光の情報は、光の強度と飛行時間であるTime-of-Flight情報を含んでおり、Transient Imageと呼ばれる。NLOSイメージングはTransient Imageから被遮蔽領域にある物体を復元する技術であるとも言える。しかし、既存手法では、結果の復元画像の品質が低いことに加え、計算量が大きいため、復元に時間を要してしまうという課題があった。そこで本研究では、スパースなデータから鮮明な画像を高速に復元することが可能なフレームワークとして知られているVision Transformerを導入したNLOSシーンの再構成手法を提案する。
Notes	会議名：AICカンファレンス2023 開催地：慶應義塾大学日吉キャンパス 日時：2023年3月4日 第2章ポスター発表要旨 ポスター要旨-11
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0037

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Vision Transformer を用いた高品質かつ高速な Non-Line-of-Sight シーン再構成手法

菅野真央 五十川麻理子

慶應義塾大学理工学部情報工学科

Abstract:

Non-Line-of-Sight (NLOS) イメージングとは、コーナー越しに遮蔽された物体の形状を復元する技術である。光を壁に照射し、その反射光が被遮蔽領域にある物体を照らす。その物体から反射した光が壁、そしてセンサへと戻ってくる。観測される光の情報は、光の強度と飛行時間である Time-of-Flight 情報を含んでおり、Transient Image と呼ばれる。NLOS イメージングは Transient Image から被遮蔽領域にある物体を復元する技術であるとも言える。しかし、既存手法では、結果の復元画像の品質が低いことに加え、計算量が大きいため、復元に時間を要してしまうという課題があった。そこで本研究では、スパースなデータから鮮明な画像を高速に復元することが可能なフレームワークとして知られている Vision Transformer を導入した NLOS シーンの再構成手法を提案する。

Keywords: Non-Line-of-Sight, Transient image, Vision transformer

1. 研究背景・目的

センシング技術は私たちの身の回りの家電製品、工場の機械、自動車など、多くの場面で使用されており、その需要は年々高まっている。しかし実際にセンシング可能な領域は、センサや観測対象物が他の物体に遮られてしまうような場合には、著しく制限されてしまうため、被遮蔽領域のセンシング技術への社会的需要も高まっている。被遮蔽領域のセンシング技術の1つとして、Non-Line-of-Sight (NLOS) イメージングが知られている。しかし、既存手法では、計測される光の信号が弱いことなどから復元を鮮明に行うことが難しく、さらに計算量が多いため NLOS イメージングの実行に時間を多く要してしまうという課題があった。そこで、本研究ではスパースなデータから高速に鮮明な画像を生成することが可能であるフレームワークとして知られている Vision Transformer を用いて復元を高品質で行う手法を提案する。

2. 方法

提案フレームワークでは、Transient Image データセットを生成し、Vision Transformer を使用した Deep Learning Model を用いて Transient Image を学習させ、高品質かつ高速な NLOS イメージングの実現を目指した。既存の Transient Image データセットでは大量のデータセットを必要とする Vision Transformer には不十分であるため、本研究では Mitsuba2-Transient-NLOS[1]を用いて Transient Image データセットの生成を行った。次に本研究で提案するネットワークについて説明する。提案ネットワークを図1に示す。

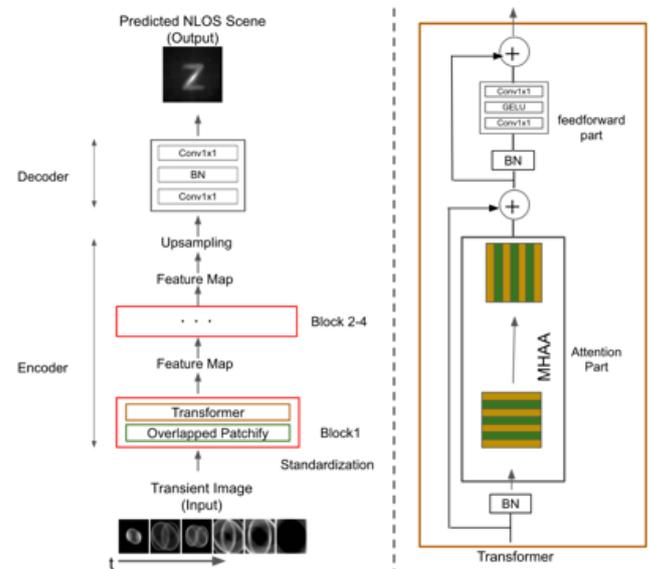


図 1: 提案ネットワーク

レンズのないカメラで撮影された画像の鮮明化を Vision Transformer を用いて行う Lensless Imaging Transformer^[2] に着想を得て、新しいフレームワークを作成した。ネットワークのエンコーダとして4つの Transformer を使用した。そして、エンコーダの後に Upsampling を行った。デコーダは畳み込み層とバッチ正規化層により構成されている。提案フレームワークの計算量のボトルネックは Transformer にある。そのため、大きな入力サイズの画像に対応するためには、Transformer の計算量を削減する必要がある。そこで、従来の Transformer で使用されている Self-Attention を、計算量を大幅に減少することが可能な Attention として知られている Axial-Attention^[3] に置き換えることで Transformer を簡略化した。

3. 結果

3.1. 定性評価

図2に実験結果を示す。本節では既存研究との実行結果の比較を定性評価により行う。既存研究は NLOS イメージング手法として知られる LCT アルゴリズムと FK アルゴリ

ズムを使用した。

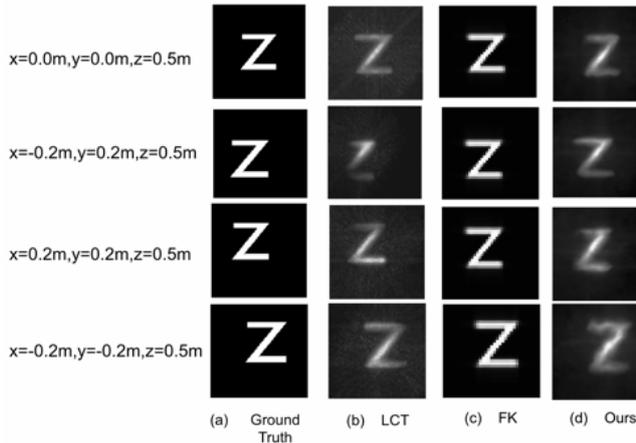


図 2: 本研究と既存研究の実行結果

図 2 を見てわかるように提案手法では、既存手法である FK アルゴリズムが最も鮮明に復元できていることがわかる。一方で、シーンによっては LCT アルゴリズムと同等の結果を得られている。

3.2. 定量評価

本研究の定量評価を、SSIM, PSNR, 実行時間により行った。SSIM とは Ground Truth と生成画像の 2 つの画像の間の輝度、コントラスト、構造の 3 つの要素を比較し、それらを乗算することで画像の類似度を算出する評価指標であり、PSNR は、画質の再現性に影響を与える、信号が取りうる最大のパワーと劣化をもたらすノイズの比率を表す評価指標である。表 1 にその結果を示す。

表 1: 実行結果の定量評価

Method	SSIM (↑)	PSNR[db] (↓)	Time[s] (↓)
LCT	0.25	14.56	0.45
FK	0.33	7.43	0.75
Ours	0.03	12.45	0.04

表 1 から見てわかるように、本研究の SSIM の結果は 0.03 となり、LCT アルゴリズムに比べ 12.0% の SSIM 値、FK アルゴリズムに比べ 9.1% の SSIM 値となり、既存研究に比べると良好でない値を示したため、画像の鮮明さにおいて精度向上を実現できなかった。一方で画質の再現性に影響を与えるノイズの比率である PSNR では本研究の結果では 12.45dB となり、FK アルゴリズムと比べ 5.02dB 劣ってしまったが、LCT アルゴリズムに比べ 2.11dB 上回る結果となった。さらに、実行時間では 1 枚あたり 0.04s となり、LCT アルゴリズムに比べ 91.0%、FK アルゴリズムに比べ 94.6% の時間で実行が可能になり、大幅な高速化を達成した。

3.3 失敗例

本研究での実験の失敗結果について説明する。復元の失敗例を図 3 にまとめる。この結果を見てわかる通り、提案手法では鮮明な Transient Image の再構成を行えなかった。さら

に、これらの Transient Image を用いて再構成を行った場合、図 3 の Ground Truth の画像の位置に物体が復元されるはずだが、図 3 では、物体が中央付近で復元されていることがわかる。

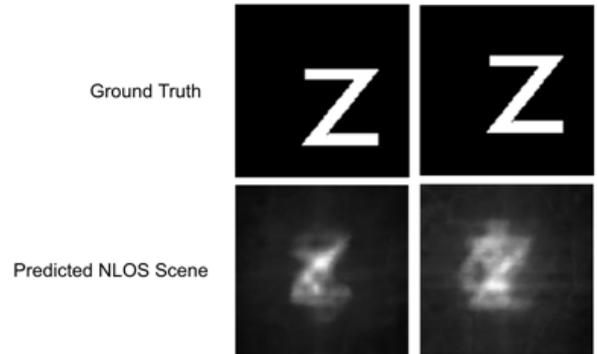


図 3: 再構成失敗例

4. 考察・結論

本手法では、既存手法に比べ、鮮明な再構成を行うことが出来なかった。それに加え、物体の位置によっては再構成を全く行えないという問題点があった。これは、データセットの Transient Image の数が少なかったこと、Transient Image データセットの中で物体の位置に偏りがあったことでこのように精度が低く、再構成が行えないような場合が発生してしまったと考えられる。

今後の課題として、精度の偏りを防ぐためにデータセットの数を充実させること、そして物体の位置の偏りもなくすことで、精度向上を図りたい。また、今回は「Z」という文字の物体のみで再構成を行ったが、他の物体での再構成、データセットに複数の物体を含んでいる場合の再構成も可能にし、実用性を高めたい。

参考文献

- [1] Diego Royo et al., Non-line-of-sight transient rendering, SIGGRAPH '22: ACM SIGGRAPH 2022 Article No.39 pp. 1-2, 2022
- [2] Xiuxi Pan et al., Image reconstruction with transformer for mask-based lensless imaging, Optics Letters Vol. 47, Issue 7, pp. 1843-1846, 2022
- [3] Huiyu Wang, Yukun Zhu, Bradley Green, Hartwig Adam, Alan L. Yuille, and Liang-Chieh Chen, Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. CoRR, Vol. abs/2003.07853, 2022