



360 度画像の切り出しに基づく 3D シーン構築の検討

石井匠¹⁾, 萬代雅希¹⁾

Takumi ISHII, and Masaki BANDAI

1) 上智大学 大学院理工学研究科（〒102-8554 東京都千代田区紀尾井町 7-1, t-ishii-4o3@eagle.sophia.ac.jp）

概要: 本研究では、360 度カメラで撮影した画像から、複数の平面画像を切り出し、それらを用いて 3D シーンを構築する手法を提案する。切り出した画像を 3D シーンの構築に利用することで、撮影回数を最小限に抑えつつ環境全体を網羅する。また、位置推定を特定の画像のみにを行い、他の画像には事前定義されたパラメータを適用することで、処理を高速化しリアルタイムでの 3D シーンの構築を実現する。

キーワード: 360 度カメラ, NeRF, 3D シーン構築, 自由視点画像生成

1. はじめに

Instant-ngp (Instant Neural Graphics Primitives) [1] は、Multiresolution Hash Encoding と小規模なニューラルネットワークを組み合わせることで、3D シーンの高速な学習とレンダリングを可能にし、VR (Virtual Reality) や AR (Augmented Reality) 等のエンターテインメント分野や、建築設計、地形測量など、幅広い分野での応用が期待されている。3D シーンの構築には、複数の視点から撮影された画像と、カメラの位置情報の推定が必要である。COLMAP[2] は、特徴点抽出・マッチング・位置推定・三次元点群の構成までを一貫して行えるオープンソースの SfM (Structure from Motion) パイプラインであり、高精度なカメラの位置推定が可能である。しかし、すべての画像に対して特徴点マッチングと最適化を行うため、リアルタイム性が損なわれる。また、屋外などの広範囲な環境で撮影をする場合、撮影にかかる時間と労力が増える課題がある。

本研究では、360 度カメラで撮影された画像から、複数の平面画像を切り出し、それらを用いてリアルタイムで 3D シーンを構築する手法を提案する。

2. 提案手法

図 1 に提案手法の処理の流れを示す。まず、360 度カメラで撮影した画像を、複数の平面画像に切り出す。次に、切り出した画像のうち、特定の画像に対して COLMAP による位置推定を行う。その他の画像には事前定義された回転パラメータを適用し、各画像に位置情報を割り当てる。最後に、切り出した画像と、対応するパラメータが記述された JSON ファイルを Instant-ngp に入力し、3D シーンを構築する。提案手法は、以下の 4 つのプロセスから構成される。

2.1 プロセス 1：平面画像の切り出し

360 度カメラを用いた撮影と画像の切り出しを行う。360 度画像を取得後、あらかじめ定義された仰角 θ と方位角 ϕ に基づき、球面投影に対する回転変換を行う。視野角を示す FoV (Field of View) を指定し、式 1 のように 45 度ず

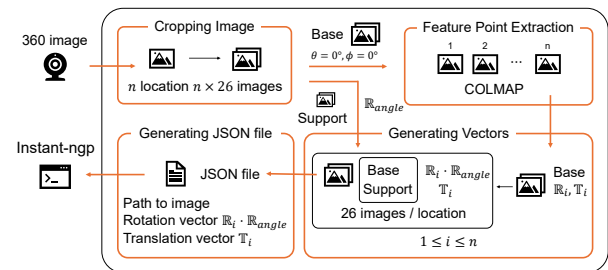


図 1: 提案手法の処理の流れ

つ変化させ、1 地点につき 26 枚の画像を切り出す。

$$angle = \begin{cases} \phi = 0^\circ & \text{if } \theta = \pm 90^\circ \\ \phi = 0^\circ, 45^\circ, 90^\circ, 135^\circ, & \text{if } \theta = 0^\circ, \pm 45^\circ \\ 180^\circ, 225^\circ, 270^\circ, 315^\circ & \end{cases} \quad (1)$$

2.2 プロセス 2：位置推定

切り出した各地点の画像の仰角 0 度、方位角 0 度の画像を基準画像、それ以外の画像を補助画像として、基準画像に対して COLMAP による位置推定を行い、 i 地点目のカメラ位置を示す並進ベクトル $\mathbf{T}_i$ と、回転を示す回転ベクトル $\mathbf{R}_i$ を推定する。

2.3 プロセス 3：ベクトル生成

位置推定を行った基準画像に基づいて、補助画像の位置ベクトルを生成する。補助画像には、基準画像と同一の並進ベクトル $\mathbf{T}_i$ を適用し、回転ベクトルは、COLMAP で得られた回転ベクトル $\mathbf{R}_i$ に、式 1 で定義される各補助画像に対応する回転パラメータ $\mathbf{R}_{angle}$ を乗じて $\mathbf{R}_i \cdot \mathbf{R}_{angle}$ とすることで決定する。

2.4 プロセス 4：JSON ファイルの生成

得られた画像や、位置情報を Instant-ngp に入力するために、JSON ファイルを生成する。JSON ファイルには焦点距離等のカメラパラメータと各画像へのパス、対応する並進ベクトルと回転ベクトルを記述し、生成した JSON ファイルと、対応する画像データを Instant-ngp に入力することで 3D シーンを構築する。

3. 性能評価

所要時間と 3D シーンの構築精度を評価する。

3.1 評価環境

評価は, Intel(R) Core(TM) i9-13900, NVIDIA GeForce RTX 4090 を搭載した PC にて行う。

単眼カメラで撮影されたすべての画像に対して COLMAP を用いて位置推定を行い, その結果をもとに Instant-ngp で 3D シーンを構築する方法を従来手法とする。提案手法では, 屋外の開けた空間において, 24 地点で 360 度カメラの Insta360 X3 を用いて撮影を行い, 各地点で 360 度画像 (11968px × 5984px) を取得した。取得した 360 度画像を, FoV を 110 度として, それぞれの地点から 26 枚の平面画像 (1920px × 1080px) に切り出し, 合計で 624 枚の画像を生成することで画像セットを得る。また, 従来手法では, 同一の環境において iPhone 12 の単眼カメラを用いて同じ視点構成で一枚ずつ手動撮影を行った画像セット (1920px × 1080px) を利用する。

3.2 リアルタイム性の評価

リアルタイム性として所要時間を評価する。所要時間は, 画像の撮影時間と画像の切り出し時間, 位置推定時間, 3D シーンの構築時間の合計時間である。なお, レンダリング時間の違いによる評価結果への影響を避けるために, 3D シーンの構築時間は 30 秒で固定する。測定結果を表 1 に示す。表 1 より所要時間は, 約 91.5%削減され, 提案手法のリアルタイム性が定量的に示された。

従来手法では, 624 枚の画像を撮影するのに 624 回の撮影が必要である。一方で, 提案手法では, 24 回の撮影で 624 枚の画像を切り出すため, 撮影にかかる時間を短縮できる。また, 位置推定においては, 従来手法では 624 枚の画像すべてに対して位置推定を行うが, 提案手法では, 各地点につき 1 枚の基準画像 (計 24 枚) のみに位置推定を行うため, 位置推定時間を短縮できる。

3.3 3D シーンの構築精度の評価

次に, 3D シーンの構築精度について評価する。PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity), LPIPS (Learned Perceptual Image Patch Similarity) を用いて, 正解画像と 3D シーン内の同地点・同視点からレンダリングされた画像を比較し, 精度を評価した。なお, 提案手法では全 624 視点, 従来手法では位置推定に成功した 597 視点について評価を行った。得られた 3D シーンを図 2, 3D シーンの構築精度の測定結果を表 2 に示す。提案手法では PSNR, SSIM の値が高く, LPIPS の値が低いことがわかる。これより, レンダリングされた画像が元の画像と類似しており, 画像構造の再現性が優れているこ

表 1: 所要時間 (秒)

手法	撮影	切り出し	推定	構築	合計
従来手法	1021	-	3011	30	4062
提案手法	210	37	69	30	346



図 2: 構築した 3D シーン (上: 従来手法 下: 提案手法)

表 2: 3D シーンの構築精度

手法	PSNR (dB)	SSIM	LPIPS
従来手法	16.93	0.372	0.720
提案手法	22.36	0.657	0.542

とが示された。また, 図 2 において, 提案手法では, 建物の壁やエッジ部分が明瞭になっており, 従来手法と比較して全体的に明瞭な 3D シーンが構築されている。

従来手法では, 特徴の乏しい画像に位置推定が機能せず, 誤差が発生する。一方で, 提案手法では, 基準画像の位置推定の結果を, 補助画像に適用することで, 特徴の乏しい画像の位置推定が間接的に可能となり, 明瞭な 3D シーンの構築ができる。

4. まとめ

本研究では, 3D シーンをリアルタイムに構築することを目的として, 360 度画像から切り出される複数の平面画像のうち, 特定の画像のみを用いた位置推定と 3D シーンの構築手法を提案した。評価より, 提案手法では, 従来手法と比較して構築までにかかる所要時間が約 91.5%削減され, リアルタイム性の向上が示された。また, PSNR などの指標においても従来手法より高い精度を示す結果が得られた。

参考文献

- [1] T. Müller *et al.*, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [2] J. L. Schönberger *et al.*, “Structure-from-Motion Revisited,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.