

3D ガウススplattingを用いた3次元再構成における静止物体の除去 Static Object Removal in 3D Reconstruction via Gaussian Splatting

菊地 佑太[†] 山田 和範[†] 小池 敦[†]
Yuta Kikuchi Kazunori Yamada Atsushi Koike

1 はじめに

環境の画像から自由視点画像を得る技術は、近年のAIの発達によって急激に発展している。3D Gaussian Splatting (3DGS) [1]はその代表的な例である。ドローンを用いて屋外撮影をする場合の固有の課題が発生するが、それらに対する先行研究もある。例えば、時刻とともに人や物の位置が変化してしまうため、3次元再構成が困難になるが(図1-a), DroneSplat[2]ではそれに対する対応が行われている(図1-b)。

本研究では映り込んだ不要物体を除去することを考える。提案手法では、テキストで除去する物体を指定することで、自由視点画像から所望の物体を除去することができる。これにより、ドローン画像の3次元再構成の活用がしやすくなる。これを実現するために、学習用画像から不要な物体をCLIP[3]等のマルチモーダル画像分類器で同定し、学習に使用する画像から削除することで(図2)、再構成させる自由視点画像の最適化を試みる。

2 先行研究

2.1 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [1]は、複数の写真から3次元再構成を行う技術であり、3次元ガウス分布を用いて点群データを補完することで、学習データの欠損に強く、滑らかに表現することができる。NeRF[4]が密度体積(voxel)表現を反復サンプリングでレンダリングするのに対し、3DGSは各ガウス・スplattingに位置・形・色・を定義して、3D空間を表現する。NeRFと比べて滑らかかつ高品質な画質が得られる上に、リアルタイムレンダリングが可能である。

2.2 DroneSplat

DroneSplat[2]は、ドローンを用いた大規模な屋外環境における動的物体への耐性を高めた3DGS [1]の拡張手法である。本手法では、Segment Anything Model v2 (SAM2) [5]を用いて各視点の学習用画像のセグメントを得る。学習過程において、動いているセグメントを段階的に学習から除外することで、動的な不要物体を考慮しない再構成が可能となる。3DGSに対し、ドローン撮影時の再構成精度および視点合成品質の両面で有意に改善されることが示されており、GPU上でのリアルタイムレンダリングも維持できる。



(a) 3DGS[1]で作成
車が動いている場面を表現できない
(b) DroneSplat[2]で生成
動体を無視して表現できる

図1: 動的シーンにおける自由視点画像生成の例

3 提案手法

ドローン画像を用いて、大規模な屋外3Dモデルを再構成することを考える。文化財保存や3Dマップの作成にあたってはみだりに駐車されている車などは景観を損ねる。また、3Dモデルを作成後に不要な物体を削除するには一定の工数を要する。

そこで本研究では、建造物や道路以外に、人やものが映り込んでいない理想的な3Dモデルを構築することを考える。DroneSplat[2]は、動いている人やものを学習用データを削除することができるが、駐車中の車など動いていない不要物体を削除することは困難である。また、ドローン撮影においては不要な人や物の写り込みを防ぐことも困難である。提案手法では、画像中から所望の人やものを指定することで、それに対応するセグメント部を削除の対象として、学習から除外する。

3.1 不要物同定アルゴリズム

ここでは、自家用車やトラック、二輪車など車に類するものを削除することを考える。まず、全ての学習用画像に対してSAM2[5]を適用し、画像中の人やものをセグメンテーションする。次に、得られたセグメントひとつひとつに対してCLIP[3]を適用し、セグメントが何であるかを予測する。「photo of a car」「photo of a truck」など車に類するリストに加え、「photo of a tree」「photo of a road sign」など画像中に写り込みそうな物のリストを用意し、CLIPを用いて、セグメントがどの説明に最も近いか予測した。予測結果が車に類するリストに入っている場合はそれを削除対象セグメントとして指定する。提案手法の全体像を図2に示す。

3.2 拡散モデルによる削除部の補完

削除対象としたセグメントを含む画像を拡散モデルを利用して補完する。図3に示すように、拡散モデルLaMa[6]を用いて学習用画像を補完する。本研究では、SAM2と

[†]東北大学 Tohoku University

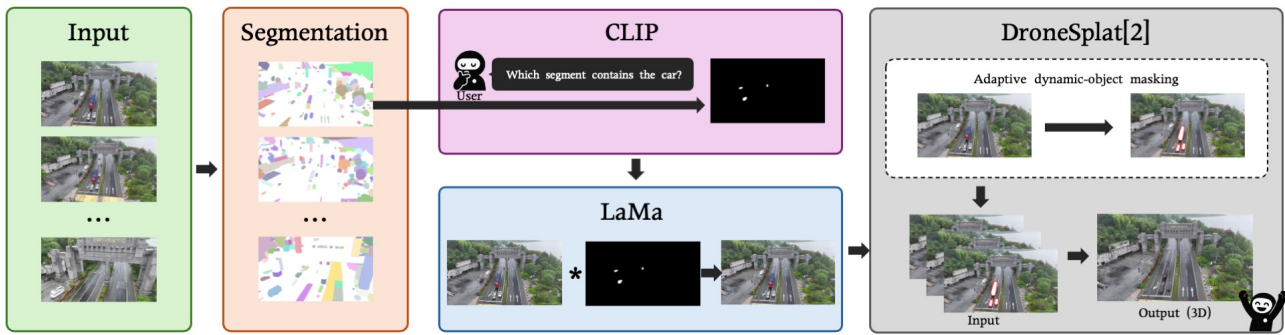


図2: 提案手法の流れ図



図3: 拡散モデルによるマスク画像の修復

LaMa による画像補完を統合したライブラリ, Inpaint Anything[7] を用いた。

4 実験

データセットは DroneSplat[2] が公開している Simingshan を用いた。24 枚のドローン画像であるが、画像中には運転中の車や停車中の車が含まれる。

5 結果

学習用画像 14 枚, テスト用画像 10 枚で評価したところ, 再構成精度は表 1 の通りであった。提案手法では DroneSplat と同等の性能を維持できている。

DroneSplat で再構成したものと提案手法で再構成したものの比較を図 4b に示す。図 4a をみると停車中の黒い車がそのまま再構成されているが, 図 4b では削除できていることがわかる。

6 おわりに

本研究ではドローン画像を用いた 3 次元再構成において, 所望の物体を削除する手法を提案した。SAM2 によるセグメンテーション処理の過程で, 検出漏れが起こることがあり, 図 4b 中の白い車のように削除しきれない場合がある。また, 拡散モデルで削除して再構成した画像を評価できる指標が存在しておらず, 定量的な評価ができていないことが課題である。今後は不要物検知のアルゴリズムを改良することで, 検知精度を向上させ, より高品質な再構成を目指す。

参考文献

- [1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023.
- [2] Jiadong Tang, Yu Gao, Dianyi Yang, Liqi Yan, Yufeng Yue, and Yi Yang. Dronesplat: 3d gaussian splatting for

表1: Simingshan データセットにおける各手法の性能比較

Method	Simingshan		
	PSNR↑	SSIM↑	LPIPS↓
NeRF[4]	19.07	0.417	0.267
3DGS[1]	19.68	0.476	0.254
DroneSplat[2]	22.76	0.759	0.152
Ours	22.35	0.744	0.174



図4: DroneSplat と提案手法の再構成画像の比較

robust 3d reconstruction from in-the-wild drone imagery, 2025.

- [3] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *CoRR*, abs/2103.00020, 2021.
- [4] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis, 2020.
- [5] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.
- [6] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. *arXiv preprint arXiv:2109.07161*, 2021.
- [7] Tao Yu, Runseng Feng, Ruoyu Feng, Jinming Liu, Xin Jin, Wenjun Zeng, and Zhibo Chen. Inpaint anything: Segment anything meets image inpainting. *arXiv preprint arXiv:2304.06790*, 2023.