Dynamic Urban Scene Modeling with 3D Gaussian Splatting from UAV Full Motion Videos

Debao Huang[1,2,3,†], Hanyang Liu[1,2,†], Ningli Xu[1,2,3], Rongjun Qin[1,2,3,4,*]

[1]Geospatial Data Analytics Laboratory, The Ohio State University, Columbus, USA
- (huang.3918, liu.12021, xu.3961, qin.324) @osu.edu
[2]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA
[3]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, USA
[4]Translational Data Analytics Institute, The Ohio State University, Columbus, USA

Keywords: Novel View Synthesis, 3D Gaussian Splatting, Unmanned Aerial Vehicle, Photogrammetry

Abstract

Reconstructing dynamic urban scenes from unmanned aerial vehicle (UAV) full-motion videos is a vital task with significant applications in urban planning, traffic analysis, and autonomous navigation. However, modeling these scenes is challenging due to their large scale and, more importantly, the ever-changing presence of dynamic objects such as vehicles and pedestrians. In recent years, emerging neural 3D scene representation approaches have gained popularity for their promising performance in novel view synthesis, and several recent works have further explored the potential of modeling large-scale and dynamic scenes. While most existing methods focus on indoor or street-level scenes, very little effort has been made to address the unique complexities of dynamic urban environments captured by UAVs. To investigate this problem, we apply a recently developed dynamic 3D Gaussian Splatting framework that decomposes urban scenes into static and dynamic elements, thereby achieving efficient and accurate modeling. We further reduce the need for auxiliary input data, thereby accommodating more general cases in which only video sequences are available. Specifically, we propose a pipeline for automatically tracking dynamic vehicles using trajectory optimization to model their natural movement, thereby eliminating the dependency on prior knowledge of vehicles — which is often unavailable in real-life scenarios. By integrating the dynamic 3D Gaussian Splatting framework with the photogrammetric reconstruction pipeline, our pipeline offers scalable and reliable 3D dynamic scene reconstruction. Our pipeline is evaluated on multiple UAV datasets, and the results demonstrate the promising quality of scene reconstruction and view synthesis.

1. Introduction

In recent decades, the development of unmanned aerial vehicles (UAVs) has facilitated various photogrammetry applications, including 3D modelling (Remondino et al., 2011; Xu et al., 2024) change detection (Andresen & Schultz-Fellenz, 2023; Xu et al., 2021), disaster monitoring (Erdelj & Natalizio, 2016), navigation (Han et al., 2022, 2024), and urban planning (Erenoglu et al., 2018; Lu et al., 2024; Muhmad Kamarulzaman et al., 2023), etc. Thanks to their high-resolution imaging, low cost, and flexible data acquisition capabilities, UAVs are particularly favored for constructing large-scale digital twins of urban scenes. The conventional photogrammetric reconstruction pipeline starts by sampling image frames from a video sequence, followed by Structure-from-Motion (SfM) to estimate the camera intrinsic and extrinsic parameters. Next, Multi-view Stereo (MVS) is performed to generate dense 3D point clouds. Often, the dense point cloud is subjected to additional post-processing, such as converting the point data into a continuous surface by generating a 3D mesh. Finally, the resulting mesh is textured using the original image data, thereby enhancing the visual fidelity and realism of the reconstructed scene. This traditional pipeline yields explicit representations — either point clouds or meshes — that are directly applicable to downstream applications. In recent years, neural radiance fields (NeRF) (Mildenhall et al., 2021) have gained popularity due to their promising performance in novel view synthesis by implicitly modeling the scene. Subsequent studies have extended these approaches to large-scale urban scenes (Tancik et al., 2022; Turki et al., 2022; Xu et al., 2024). Meanwhile, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) represents another approach that explicitly models scenes as collections of 3D Gaussians, enabling efficient rendering.

While most existing efforts have concentrated on modeling static scenes, relatively few studies have tackled the challenges of dynamic scene modeling. Full-motion videos captured by UAVs not only include static objects, such as buildings and roads, but also record the movement of dynamic objects, such as vehicles and pedestrians. The traditional photogrammetric reconstruction pipeline inherently fails to model dynamic objects because they do not produce consistent feature correspondences between frames, which leads to errors in camera pose estimation and ultimately in 3D reconstruction. In the context of NeRF and 3DGS, several studies have attempted to model dynamic scenes by decomposing complex scenes into distinct dynamic and static components. However, these approaches are typically tailored to indoor and object-level scenes (G. Wu et al., 2024) or street view ground sequences (Fischer et al., 2024), while studies on UAV datasets remain underexplored. Unlike ground-level data collection (Geiger et al., 2012; Han & Yilmaz, 2021, 2022; Liao et al., 2023), the perspective and resolution of UAV data present additional challenges for modeling moving vehicles, as their accurate representation largely depends on the quality of the reconstruction of the static environment. The absence of oblique views also hinders the rendering of ground scenes, as these perspectives typically fall outside the distribution of views

[†] Equal contribution
[*] Corresponding author, 2036 Neil Avenue, Columbus, Ohio, USA. qin.324@osu.edu

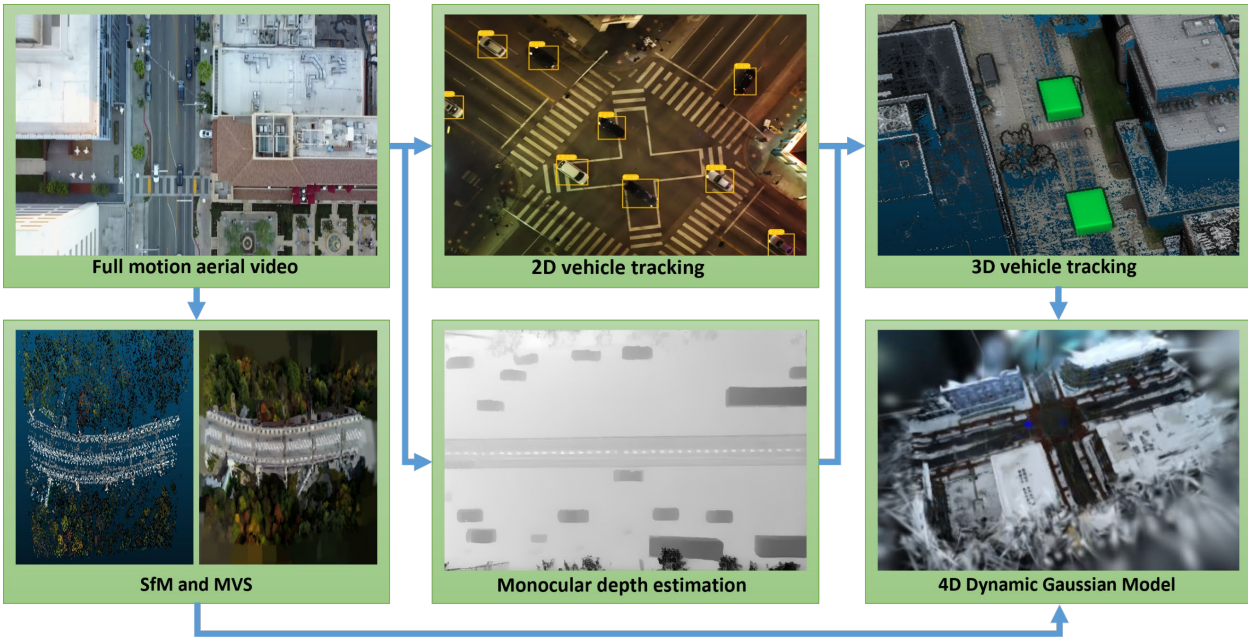

Figure 1. Overview of the proposed pipeline. Beginning with the full-motion aerial video (top-left), the framework performs 2D vehicle detection and tracking (top-center) and monocular depth estimation (bottom-center). Their outputs are combined to derive refined 3D trajectories (top-right). Meanwhile, structure-from-motion and multi-view stereo (bottom-left) are used to generate a dense reconstruction of the static environment. Finally, the static and dynamic components are integrated into a 4D dynamic Gaussian representation (bottom-right).

encountered during model training. In addition, most existing works assume prior knowledge of camera calibration and poses, the 3D point clouds of dynamic objects (typically provided by LiDAR), and the 3D bounding boxes of these objects. Although the availability of auxiliary data can facilitate network training and yield realistic novel view synthesis, such prior information is often unavailable in real-life scenarios.

In this work, we aim to assess the recently emerged dynamic 3DGS framework for its feasibility in modeling dynamic urban scenes from UAV full-motion videos. We integrate the conventional photogrammetric reconstruction pipeline with the recent dynamic 3DGS framework to effectively address the challenges of dynamic urban scene modeling. We adapt a recent approach (Fischer et al., 2024) and propose modifications based on it to further enhance dynamic scene modeling on UAV datasets without auxiliary data. Specifically, we implement an automated pipeline that integrates photogrammetric reconstruction for initializing the static scene, monocular depth estimation for initializing dynamic objects, object tracking to obtain the 3D bounding boxes of dynamic objects across consecutive frames, and a dynamic 3DGS framework that takes the processed data as input to model the dynamic urban scene. The remainder of this paper is structured as follows: Section 2 presents a comprehensive review of recent works on dynamic scene modeling; Section 3 details our proposed pipeline, which combines photogrammetric reconstruction with a dynamic 3DGS framework; Section 4 presents the experimental results along with their analysis; and Section 5 concludes the paper by offering our insights into future work.

2. Related Work

2.1 Scene Representations

Scene representations lie at the crossroads of computer vision and graphics, providing powerful tools for numerous downstream tasks including view synthesis (Gao et al., 2024; Mildenhall et al., 2021; Müller et al., 2022), SfM (Huang et al., 2022, 2024; Moulon et al., 2017; Schonberger & Frahm, 2016; Suh & Ouimet, 2023), dense matching (Huang & Qin, 2023), and 3D registration (Tao et al., 2023; Xu et al., 2023; Xu & Qin, 2024). For decades, researchers have tackled this challenge across diverse settings—including forward-looking scenes (Mildenhall et al., 2021; Müller et al., 2022), indoor scenes, street scenes (Tancik et al., 2022; Xie et al., 2023), UAV scenes (Turki et al., 2022; J. Wu et al., 2023; Xu et al., 2024; F. Zhou et al., 2024), and satellite scenes (Derksen & Izzo, 2021; Fu et al., 2023; Sariturk et al., 2023; Wang et al., 2024). Broadly, scene modeling techniques fall into two categories: implicit and explicit representations. For example, NeRF (Mildenhall et al., 2021) employs an implicit approach by using a multilayer perceptron (MLP) that takes a 3D location and viewing direction as input to produce appearance and opacity values. This method has achieved remarkable photorealism on small-scale, indoor datasets. In contrast, the recently proposed Gaussian Splatting (Kerbl et al., 2023) adopts an explicit strategy by representing scenes as a collection of 3D Gaussians, which enables not only faster rendering but also superior performance on large-scale outdoor datasets.

2.2 Dynamic Scene Modeling

Dynamic scene modeling has recently garnered significant attention. In implicit approaches, the parametric function is augmented to incorporate a temporal dimension (Liu et al., 2023; Park et al., 2021; Pumarola et al., 2021). However, achieving satisfactory results with these methods typically requires datasets that provide multi-view data for each time frame—a requirement that is often met only in indoor scenes (Pumarola et al., 2021). Conversely, explicit methods capture dynamics by modeling particle-level motions, such as scene flow (Li et al., 2021; Xian et al., 2021) and by representing the rigid transformations of local geometric primitives (Luiten et al., 2024). Traditional methods, in contrast, focus on decomposing scenes into high-level elements and representing entities along with their spatial

relationships through directed graphs (Salas-Moreno et al., 2013; Tulsiani et al., 2018).

2.3 Dynamic Urban Scene Modeling

The modeling of dynamic urban scene presents significant challenges due to the large scale of scenes, frequent occlusions, and the complex nature of dynamic objects. Some approaches require additional depth priors—such as LiDAR—to provide supplementary information like camera exposure (Martin-Brualla et al., 2021; Rematas et al., 2022; Xie et al., 2023). Other studies have tackled the problem in ground-level settings by decomposing scenes into dynamic and static components and modeling each separately (Fischer et al., 2024; X. Zhou et al., 2024). However, most existing methods rely on datasets from autonomous driving, leaving the modeling of dynamic scenes from UAV full-motion videos relatively unexplored. Moreover, the reliance on auxiliary data such as LiDAR point clouds can be burdensome for general users.

3. Methodology

We present a complete pipeline for reconstructing urban 3D environments from a single full-motion video while accurately modeling moving objects. As shown in Figure 1, our approach integrates scene reconstruction, vehicle tracking, refined depth estimation, and 4D Gaussian splatting to deliver a dynamic representation of the urban landscape.

3.1 Photogrammetric Reconstruction

To accurately capture the urban environment, we extract frames from the full-motion video at a rate of one frame per second. This rate is chosen to balance computational efficiency with the need for robust vehicle tracking. The selected frames serve as inputs for structure-from-motion (SfM) (Schonberger & Frahm, 2016) which estimates camera poses and generates a sparse point cloud that captures the essential geometry of the scene. Building on this, multi-view stereo (MVS) (Schönberger et al., 2016) is employed to create a dense reconstruction, producing a detailed point cloud that reflects the intricate structure of the urban landscape. This dense point cloud forms the foundational model for the static components of the scene and is subsequently used to initialize the Gaussian field representation.

3.2 Dynamic Vehicle Tracking

To reconstruct dynamic vehicles, it is necessary to determine each vehicle's position for every video frame. We begin by detecting the two-dimensional trajectory of each vehicle in the frames. Next, we back-project these 2D trajectories into three-dimensional space to obtain their precise spatial locations.

2D Tracking. We first detect vehicles in each video frame using a YOLOv8 object detection network (Varghese & M., 2024) trained on the VisDrone2018 dataset (Zhu et al., 2022). Then DeepSORT is used to link these detections into continuous 2D trajectories, filling in any missing data through linear interpolation.

3D Tracking. To extend these trajectories into three dimensions, we estimate the depth of moving vehicles. Given dense reconstruction fails to capture the depth of moving objects, we use a monocular depth estimation method, Depth Pro (Bochkovskii et al., 2024) to generate approximate depth maps. We then refine these estimates by aligning them with depth data from the dense reconstruction. RANSAC is used to filter out outliers in the process. Finally, we project the 2D trajectories into 3D space.

Dataset	PSNR	SSIM	LPIPS
Campus region	25.93	0.84	0.13
Night crossroad	24.62	0.71	0.21
Viaduct region	27.88	0.90	0.11

Table 1. Statistical results of reconstruction quality for the three datasets.

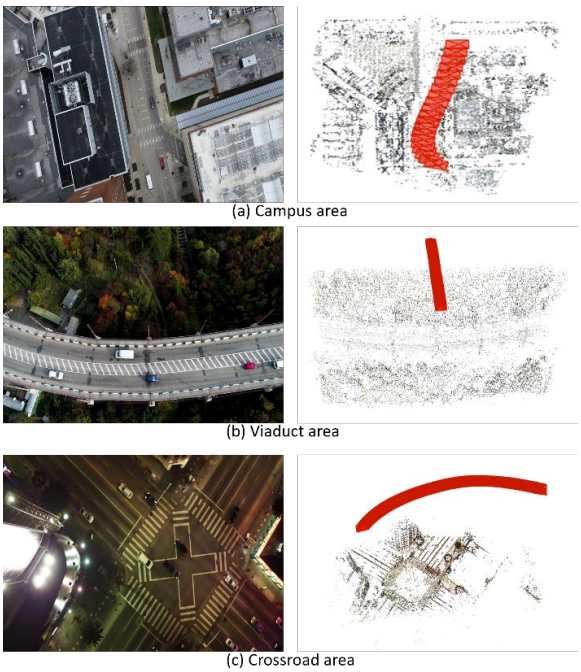

Figure 2. Overview of the dataset. Left: Sample image. Right: Sparse reconstruction results.

Trajectory optimization. Errors in detection, interpolation, and depth estimation can lead to inaccuracies in the 3D tracks. To address this, we impose a smoothness constraint that reflects the natural movement of vehicles. We adjust the vehicle's 2D centers and depths, re-project them into 3D space, and apply cubic spline interpolation to produce smooth trajectories. The deviation between these refined paths and the original estimates is minimized during optimization based on Equation 1:

$$\mathcal{L} = \|P(p_{2D} + \Delta p, d + \Delta h) - S(P(p_{2D} + \Delta p, d) + \Delta h)\|_2 \quad (1)$$

In the formulation, p_{2D} and d represent the original 2D center and depth, respectively; Δp and Δh are the adjustments applied to correct detection and depth errors; P(·) re-projects these adjusted values into 3D space; and S(·) generates the smoothed trajectory via cubic spline interpolation.

3.3 Urban Dynamic Reconstruction

We deploy a modified 4DGF model to represent the static urban scene and dynamic vehicles separately.

Static scene representation. Instead of relying on lidar scans as in the original 4DGF, we use the dense point cloud from the photogrammetry stage to initialize the Gaussians. Each Gaussian is parameterized by its position, scale, and spherical harmonics. A neural network is used to predict spherical harmonics instead

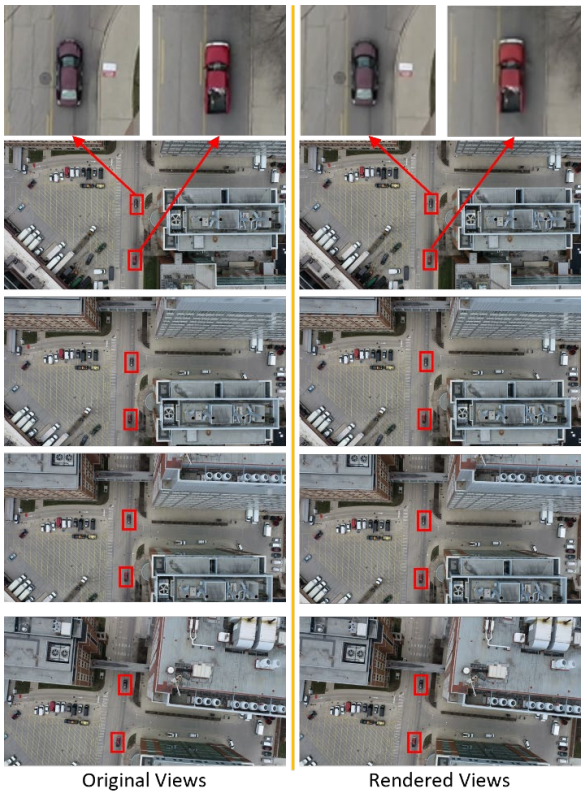

Figure 3. Visual comparison between original views and rendered views using the same camera poses. The rendering quality is promising for both static scenes (buildings) and dynamic scenes (moving vehicles) outlined in red rectangles.

of storing them explicitly, thereby reducing memory usage for large-scale reconstructions.

Dynamic vehicle representation. Each vehicle is modeled in a canonical space defined by its length, width, and height. We reconstruct the 3D shape of each vehicle by projecting its 2D pixels into 3D space, and then PCA is used to determine its dimensions. The resulting 3D points serve as the initial positions for the Gaussians representing the vehicles.

Rendering. Each vehicle is transformed from its canonical space back into the original 3D scene based on the center position estimated in each frame. The vehicle’s orientation is computed using its 3D tracking forward direction and further refined with a learnable adjustment. Finally, the static scene and dynamic vehicles are merged, and the Gaussians are optimized by minimizing the difference between the rendered image and the corresponding video frame as shown in Equation 2:

$$\mathcal{L} = \lambda_{rgb}\mathcal{L}_{rgb}(\hat{I}, I) + \lambda_{ssim}L_{ssim}(\hat{I}, I) \quad (2)$$

where $\mathcal{L}_{rgb}$ is the L_1 norm measuring pixel-wise differences between the rendered image $\hat{I}$ and the ground truth image I, and L_{ssim} is the structural similarity index measure.

4. Experimental Results

4.1 Datasets and Metrics

We evaluate our pipeline using three distinct UAV full-motion videos, as shown in Figure 2. The first video shows a campus area

Figure 4. Novel oblique viewpoint synthesized from the top-down campus dataset, highlighting structural details such as building facades and vehicle contours that are difficult to observe from the original perspective.

Figure 5. A stationary and global viewpoint synthesized from the viaduct footage, revealing moving vehicles more clearly and showcasing the effectiveness of the proposed pipeline.

from a top-down perspective (Ground Sample Distance = 0.04 m), the second captures a viaduct region transitioning from top-down to a side view (GSD = 0.05m), and the third depicts a night-time crossroad from a similar vantage (GSD = 0.07m). We assess the quality of our rendered images using standard metrics PSNR, SSIM (Hore & Ziou, 2010) and LPIPS (Zhang et al., 2018), following the protocol of (Fischer et al., 2024).

4.2 Implementation Details

Sparse and dense reconstructions are performed using COLMAP (Schonberger & Frahm, 2016) to generate the camera pose and

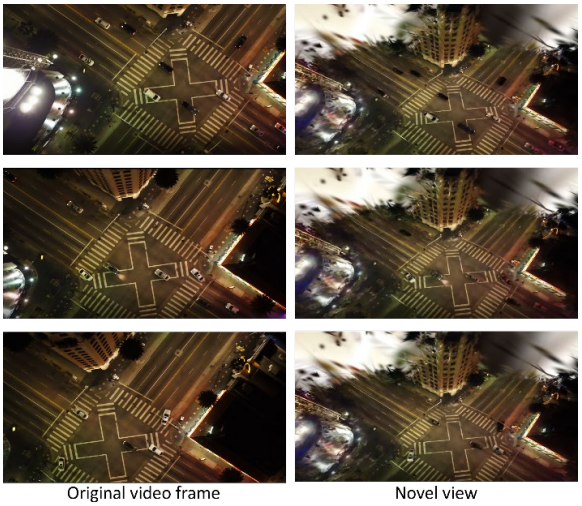

Figure 6 Novel oblique viewpoint synthesized from the top-down night crossroad dataset, underscoring the robustness of the proposed pipeline in challenging low-light conditions.

dense point cloud as the initial of the Gaussians. For 2D vehicle tracking, we employ a pre-trained DeepSORT model. In the dynamic reconstruction rendering stage, λrgb and λssim are set to 0.8 and 0.2, respectively.

4.3 Results

To evaluate our pipeline, we render the same viewpoints as those in the original UAV videos and compare the reconstructed images with the ground-truth frames. As summarized in Table 1, our pipeline performs consistently well across the three datasets, achieving high PSNR and SSIM values alongside low LPIPS scores, consistent with the original performance of 4DGF on the street scenes. Figure 3 presents a visual comparison of the campus dataset, showcasing the strong alignment between rendered images and real-world views. Notably, dynamic objects are well-aligned with the original images, highlighting the accuracy of our vehicle tracking in detecting vehicle locations in each frame. These observations prove the quantitative metrics reported in Table 1.

To demonstrate the flexibility of our approach, we generate novel view videos in which the angles were not part of the original input. Figure 4 shows a tilted viewpoint of the campus region, exposing side details of buildings and vehicles that remain hidden in the original top-down shots. However, the rendered image exhibits slightly lower quality than the original view due to the Gaussian color being predicted based on the viewpoint angle. Since the original view lacks this perspective data, rendering from a new angle results in moderate quality degradation. Figure 5 compares the original viaduct footage — which captures only partial views of the scene — to frames rendered from a stationary, global perspective that provides a broader field of view and a clearer depiction of overall dynamic changes. This vantage may prove advantageous for transportation analyses by enabling a more comprehensive understanding of vehicle motion. Notably, some noisy Gaussians appear along the edges of the rendered videos, arising from missing data in regions beyond the coverage of the original full-motion video. Lastly, Figure 6 highlights our pipeline's robustness in a nighttime crossroad scenario, where it accurately depicts multiple vehicles—including turning ones—even in low-light conditions. This demonstrates the adaptability and reliability of our pipeline across diverse urban environments.

5. Conclusion

In this work, we assess the feasibility of the recently developed dynamic 3DGS framework for modeling dynamic urban scenes from UAV full-motion videos. We apply a pipeline that incorporates both a photogrammetric reconstruction framework and a recent dynamic 3DGS framework to model dynamic urban scenes from UAV full-motion videos. We propose several modules that derive the necessary information for the dynamic 3DGS framework directly from video data, thereby eliminating the need for auxiliary data, which is often unavailable to typical users. Our proposed pipeline is evaluated on multiple UAV datasets, and the qualitative and quantitative results demonstrate its effectiveness. This indicates the potential of our pipeline for broader applications by scaling up scene sizes and accommodating more complex scenarios. Future work could focus on improving the tracking of moving objects across frames to achieve smoother and more realistic rendering in synthetic videos. In addition, improving the rendering quality of different angles by incorporating the recent diffusion models is also worth exploring.

Acknowledgements

This work was partially sponsored by the United States Air Force Research Laboratory and the United States AFRL Regional Hub and was accomplished under Cooperative Agreement Number FA8750-22-2-0501. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. This work was also partially supported by the Office of Naval Research (ONR, Award No. N00014-20-1-2141 and N00014-23-1-2670).

References

Andresen, C. G., & Schultz-Fellenz, E. S. (2023). Change Detection Applications in the Earth Sciences Using UAS-Based Sensing: A Review and Future Opportunities. *Drones*, *7*(4), Article 4. https://doi.org/10.3390/drones7040258

Bochkovskii, A., Delaunoy, A., Germain, H., Santos, M., Zhou, Y., Richter, S. R., & Koltun, V. (2024). *Depth Pro: Sharp Monocular Metric Depth in Less Than a Second* (arXiv:2410.02073). arXiv. https://doi.org/10.48550/arXiv.2410.02073

Derksen, D., & Izzo, D. (2021). *Shadow neural radiance fields for multi-view satellite photogrammetry*. https://openaccess.thecvf.com/content/CVPR2021W/EarthVision/html/Derksen_Shadow_Neural_Radiance_Fields_for_Multi-View_Satellite_Photogrammetry_CVPRW_2021_paper.html

Erdelj, M., & Natalizio, E. (2016). UAV-assisted disaster management: Applications and open issues. *2016 International Conference on Computing, Networking and Communications (ICNC)*, 1–5. https://doi.org/10.1109/ICCNC.2016.7440563

Erenoglu, R. C., Erenoglu, O., & Arslan, N. (2018). Accuracy Assessment of Low Cost UAV Based City Modelling for Urban

Planning. *Tehnički Vjesnik*, *25*(6), 1708–1714. https://doi.org/10.17559/TV-20170904202055

Fischer, T., Kulhanek, J., Bulò, S. R., Porzi, L., Pollefeys, M., & Kontschieder, P. (2024). *Dynamic 3D Gaussian Fields for Urban Areas* (arXiv:2406.03175). arXiv. https://doi.org/10.48550/arXiv.2406.03175

Fu, Q., Tong, X., Liu, S., Ye, Z., Jin, Y., Wang, H., & Hong, Z. (2023). GPU-Accelerated PCG Method for the Block Adjustment of Large-Scale High-Resolution Optical Satellite Imagery Without GCPs. *Photogrammetric Engineering & Remote Sensing*, *89*(4), 211–220. https://doi.org/10.14358/PERS.22-00051R2

Gao, Z., Teng, W., Chen, G., Wu, J., Xu, N., Qin, R., Feng, A., & Zhao, Y. (2024). *Skyeyes: Ground Roaming using Aerial View Images* (arXiv:2409.16685). arXiv. https://doi.org/10.48550/arXiv.2409.16685

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

Han, Y., Toth, C., & Yilmaz, A. (2024). *UAS Visual Navigation in Large and Unseen Environments via a Meta Agent*. https://isprs-annals.copernicus.org/articles/X-2-2024/105/2024/isprs-annals-X-2-2024-105-2024.html

Han, Y., Wei, J., & Yilmaz, A. (2022). UAS Navigation in the Real World Using Visual Observation. *2022 IEEE Sensors*, 1–4. https://doi.org/10.1109/SENSORS52175.2022.9967103

Han, Y., & Yilmaz, A. (2021). *DYNAMIC ROUTING FOR NAVIGATION IN CHANGING UNKNOWN MAPS USING DEEP REINFORCEMENT LEARNING*. https://www.proquest.com/openview/04d25bedbbd978c234882d9fb2f6f351/1?pq-origsite=gscholar&cbl=2037681

Han, Y., & Yilmaz, A. (2022). Learning to Drive Using Sparse Imitation Reinforcement Learning. *2022 26th International Conference on Pattern Recognition (ICPR)*, 3736–3742. https://doi.org/10.1109/ICPR56361.2022.9956121

Hore, A., & Ziou, D. (2010). *Image Quality Metrics: PSNR vs. SSIM*. https://ieeexplore.ieee.org/abstract/document/5596999

Huang, D., Elhashash, M., & Qin, R. (2022). Constrained Bundle Adjustment for Structure From Motion Using Uncalibrated Multi-Camera Systems. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *V-2–2022*, 17–22. https://doi.org/10.5194/isprs-annals-v-2-2022-17-2022

Huang, D., & Qin, R. (2023). A critical analysis of internal reliability for uncertainty quantification of dense image matching in multi-view stereo. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-1/W1-2023*, 517–524. https://doi.org/10.5194/isprs-annals-X-1-W1-2023-517-2023

Huang, D., Qin, R., & Elhashash, M. (2024). Bundle adjustment with motion constraints for uncalibrated multi-camera systems at the ground level. *ISPRS Journal of Photogrammetry and Remote Sensing*, *211*, 452–464. https://doi.org/10.1016/j.isprsjprs.2024.04.023

Kerbl, B., Kopanas, G., Leimkuehler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, *42*(4), 1–14. https://doi.org/10.1145/3592433

Li, Z., Niklaus, S., Snavely, N., & Wang, O. (2021). *Neural scene flow fields for space-time view synthesis of dynamic scenes*. https://openaccess.thecvf.com/content/CVPR2021/html/Li_Neural_Scene_Flow_Fields_for_Space-Time_View_Synthesis_of_Dynamic_CVPR_2021_paper.html

Liao, Y., Xie, J., & Geiger, A. (2023). KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(3), 3292–3310. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2022.3179507

Liu, Y.-L., Gao, C., Meuleman, A., Tseng, H.-Y., Saraf, A., Kim, C., Chuang, Y.-Y., Kopf, J., & Huang, J.-B. (2023). *Robust dynamic radiance fields*. https://openaccess.thecvf.com/content/CVPR2023/html/Liu_Robust_Dynamic_Radiance_Fields_CVPR_2023_paper.html

Lu, J., Li, J., Yu, F. R., Jiang, W., & Feng, W. (2024). UAV-Assisted Heterogeneous Cloud Radio Access Network With Comprehensive Interference Management. *IEEE Transactions on Vehicular Technology*, *73*(1), 843–859. IEEE Transactions on Vehicular Technology. https://doi.org/10.1109/TVT.2023.3306359

Luiten, J., Kopanas, G., Leibe, B., & Ramanan, D. (2024). Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. *2024 International Conference on 3D Vision (3DV)*, 800–809. https://doi.org/10.1109/3DV62453.2024.00044

Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). *Nerf in the wild: Neural radiance fields for unconstrained photo collections*. https://openaccess.thecvf.com/content/CVPR2021/html/Martin-Brualla_NeRF_in_the_Wild_Neural_Radiance_Fields_for_Unconstrained_Photo_CVPR_2021_paper.html?ref=labelbox.ghost.io

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, *65*(1), 99–106. https://doi.org/10.1145/3503250

Moulon, P., Monasse, P., Perrot, R., & Marlet, R. (2017). OpenMVG: Open Multiple View Geometry. In B. Kerautret, M. Colom, & P. Monasse (Eds.), *Reproducible Research in Pattern Recognition* (pp. 60–74). Springer International Publishing. https://doi.org/10.1007/978-3-319-56414-2_5

Muhmad Kamarulzaman, A. M., Wan Mohd Jaafar, W. S., Mohd Said, M. N., Saad, S. N. M., & Mohan, M. (2023). UAV Implementations in Urban Planning and Related Sectors of Rapidly Developing Nations: A Review and Future Perspectives for Malaysia. *Remote Sensing*, *15*(11), Article 11. https://doi.org/10.3390/rs15112845

Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, *41*(4), 1–15. https://doi.org/10.1145/3528223.3530127

Park, K., Sinha, U., Barron, J. T., Bouaziz, S., Goldman, D. B., Seitz, S. M., & Martin-Brualla, R. (2021). *Nerfies: Deformable neural radiance fields*. https://openaccess.thecvf.com/content/ICCV2021/html/Park_Nerfies_Deformable_Neural_Radiance_Fields_ICCV_2021_paper.html

Pumarola, A., Corona, E., Pons-Moll, G., & Moreno-Noguer, F. (2021). *D-nerf: Neural radiance fields for dynamic scenes*. https://openaccess.thecvf.com/content/CVPR2021/html/Pumarola_D-NeRF_Neural_Radiance_Fields_for_Dynamic_Scenes_CVPR_2021_paper.html?ref=labelbox.ghost.io

Rematas, K., Liu, A., Srinivasan, P. P., Barron, J. T., Tagliasacchi, A., Funkhouser, T., & Ferrari, V. (2022). *Urban radiance fields*. https://openaccess.thecvf.com/content/CVPR2022/html/Rematas_Urban_Radiance_Fields_CVPR_2022_paper.html

Remondino, F., Barazzetti, L., Nex, F. C., Scaioni, M., & Sarazzi, D. (2011). *UAV photogrammetry for mapping and 3D modeling: Current status and future perspectives*. https://research.utwente.nl/en/publications/uav-photogrammetry-for-mapping-and-3d-modeling-current-status-and

Salas-Moreno, R. F., Newcombe, R. A., Strasdat, H., Kelly, P. H. J., & Davison, A. J. (2013). *Slam++: Simultaneous localisation and mapping at the level of objects*. https://openaccess.thecvf.com/content_cvpr_2013/html/Salas-Moreno_SLAM_Simultaneous_Localisation_2013_CVPR_paper.html

Sariturk, B., Kumbasar, D., & Seker, D. Z. (2023). Comparative Analysis of Different CNN Models for Building Segmentation from Satellite and UAV Images. *Photogrammetric Engineering & Remote Sensing*, *89*(2), 97–105. https://doi.org/10.14358/PERS.22-00084R2

Schonberger, J. L., & Frahm, J.-M. (2016). Structure-From-Motion Revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4104–4113.

Schönberger, J. L., Zheng, E., Frahm, J.-M., & Pollefeys, M. (2016). Pixelwise View Selection for Unstructured Multi-View Stereo. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 501–518). Springer International Publishing. https://doi.org/10.1007/978-3-319-46487-9_31

Suh, J. W., & Ouimet, W. (2023). Generation of High-Resolution Orthomosaics from Historical Aerial Photographs Using Structure-from-Motion and Lidar Data. *Photogrammetric Engineering & Remote Sensing*, *89*(1), 37–46. https://doi.org/10.14358/PERS.22-00063R2

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B. P., Srinivasan, P., Barron, J. T., & Kretzschmar, H. (2022). Block-NeRF: Scalable Large Scene Neural View Synthesis. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8238–8248. https://doi.org/10.1109/CVPR52688.2022.00807

Tao, W., Xu, D., Chen, X., & Tan, G. (2023). A Powerful Correspondence Selection Method for Point Cloud Registration Based on Machine Learning. *Photogrammetric Engineering & Remote Sensing*, *89*(11), 703–712. https://doi.org/10.14358/PERS.23-00046R2

Tulsiani, S., Gupta, S., Fouhey, D. F., Efros, A. A., & Malik, J. (2018). *Factoring shape, pose, and layout from the 2d image of a 3d scene*. https://openaccess.thecvf.com/content_cvpr_2018/html/Tulsiani_Factoring_Shape_Pose_CVPR_2018_paper.html

Turki, H., Ramanan, D., & Satyanarayanan, M. (2022). *Mega-NERF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs*. 12922–12931. https://doi.org/10.1109/cvpr52688.2022.01258

Varghese, R., & M., S. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 1–6. https://doi.org/10.1109/ADICS58448.2024.10533619

Wang, J., Yang, D., Xie, Z., Wang, H., Hao, Z., Zhou, F., & Wang, X. (2024). Research Progress of Optical Satellite Remote Sensing Monitoring Asphalt Pavement Aging. *Photogrammetric Engineering & Remote Sensing*, *90*(8), 471–482. https://doi.org/10.14358/PERS.23-00045R2

Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., & Wang, X. (2024). *4d gaussian splatting for real-time dynamic scene rendering*. https://openaccess.thecvf.com/content/CVPR2024/html/Wu_4D_Gaussian_Splatting_for_Real-Time_Dynamic_Scene_Rendering_CVPR_2024_paper.html

Wu, J., Fu, S., Chen, P., Chen, Q., & Pan, X. (2023). Validation of Island 3D-mapping Based on UAV Spatial Point Cloud Optimization: A Case Study in Dongluo Island of China. *Photogrammetric Engineering & Remote Sensing*, *89*(3), 173–182. https://doi.org/10.14358/PERS.22-00109R2

Xian, W., Huang, J.-B., Kopf, J., & Kim, C. (2021). *Space-time neural irradiance fields for free-viewpoint video*. https://openaccess.thecvf.com/content/CVPR2021/html/Xian_Space-Time_Neural_Irradiance_Fields_for_Free-Viewpoint_Video_CVPR_2021_paper.html

Xie, Z., Zhang, J., Li, W., Zhang, F., & Zhang, L. (2023). *S-NeRF: Neural Radiance Fields for Street Views* (arXiv:2303.00749). arXiv. https://doi.org/10.48550/arXiv.2303.00749

Xu, N., Huang, D., Song, S., Ling, X., Strasbaugh, C., Yilmaz, A., Sezen, H., & Qin, R. (2021). A volumetric change detection framework using UAV oblique photogrammetry – a case study of ultra-high-resolution monitoring of progressive building collapse. *International Journal of Digital Earth*, *14*(11), 1705–1720. https://doi.org/10.1080/17538947.2021.1966527

Xu, N., & Qin, R. (2024). Large-scale DSM registration via motion averaging. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-1–2024*, 275–282. https://doi.org/10.5194/isprs-annals-x-1-2024-275-2024

Xu, N., Qin, R., Huang, D., & Remondino, F. (2024). Multi-tiling neural radiance field (NeRF)—Geometric assessment on large-scale aerial datasets. *The Photogrammetric Record*, *39*(188), 718–740. https://doi.org/10.1111/phor.12498

Xu, N., Qin, R., & Song, S. (2023). Point cloud registration for LiDAR and photogrammetric data: A critical synthesis and performance analysis on classic and deep learning algorithms. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *8*, 100032. https://doi.org/10.1016/j.ophoto.2023.100032

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*. 586–595. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html

Zhou, F., Liu, L., Hu, H., Jin, W., Zheng, Z., Li, Z., Ma, Y., & Wang, Q. (2024). An Improved YOLO Network for Insulator and Insulator Defect Detection in UAV Images. *Photogrammetric Engineering & Remote Sensing*, *90*(6), 355–361. https://doi.org/10.14358/PERS.23-00074R2

Zhou, X., Lin, Z., Shan, X., Wang, Y., Sun, D., & Yang, M.-H. (2024). *Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes*. https://openaccess.thecvf.com/content/CVPR2024/html/Zhou_DrivingGaussian_Composite_Gaussian_Splatting_for_Surrounding_Dynamic_Autonomous_Driving_Scenes_CVPR_2024_paper.html

Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H. (2022). Detection and Tracking Meet Drones Challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7380–7399. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2021.3119563