IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641410

SfM-MVS Photogrammetry with UAS: Leveraging Image Segmentation for Efficient Mapping in Dynamic Coastal Zones

Mohammad Pashaei, Michael J. Starek, Jacob Berryhill, and José Pilartes-Congo

***Abstract*—Structure from Motion (SfM) photogrammetry, in conjunction with the Multi-View Stereo (MVS) technique, collectively known as SfM-MVS, emerges as a cost-effective solution for reconstructing 3D structures in real-world environments through the utilization of overlapping images. While SfM photogrammetry finds widespread use in remote sensing applications, challenges persist regarding reconstruction quality, scene segmentation, and computational complexity and efficiency. This paper introduces a workflow wherein semantically segmented images guide the SfM-MVS processing of overlapping images for reconstruction. The proposed workflow is applied to address two challenging tasks. The first task focuses on reconstructing a narrow pier situated over dynamic open ocean waves, while the second task involves simultaneous reconstruction and scene (point cloud) segmentation. Semantic labels assigned to pixels play a crucial role in determining the inclusion or exclusion of specific pixel sets during SfM-MVS processing. This experimental study underscores the promising potential of the proposed workflow to seamlessly integrate with the conventional SfM-MVS processing workflow. The approach not only augments the reconstruction quality in challenging environments but also advances the level of automation in generating spatial products within established SfM photogrammetry software suites. These findings contribute to the ongoing discourse on improving SfM-MVS methodologies for enhanced reconstruction outcomes and increased efficiency in spatial product generation.**

I. Introduction

Structure-from-motion (SfM) photogrammetry is a highly efficient alternative to traditional digital photogrammetry, used to extract the geometric structure of objects or environments with intricate detail from a sequence of overlapping images. In recent years, small uncrewed aircraft systems (UASs) equipped with digital cameras have become effective tools for capturing aerial imagery in various remote sensing (RS) applications. [1].

The main products of SfM computations include the geometry of the reconstructed scene, the position and orientation of the camera at each exposure station (the camera's Exterior Orientation or EO parameters), the internal geometry of the camera (the camera's Interior Orientation or IO parameters), and a sparse point cloud representing the 3D structure of the object or surveyed area. SfM processing is commonly paired with multi-view stereo (MVS) algorithms to generate a dense point cloud that intricately captures the details of the study area.

The quality of the SfM solution depends on several factors, notably the quality of the detected keypoints and generated tie points in the processed image set. Challenges arise with poor surface texture or repetitive patterns, which introduce higher uncertainty and noise into the SfM solution. Additionally, SfM computations may encounter difficulties when a significant portion of pixels in the image set depicts dynamic surfaces, such as moving water, potentially leading to processing failures.

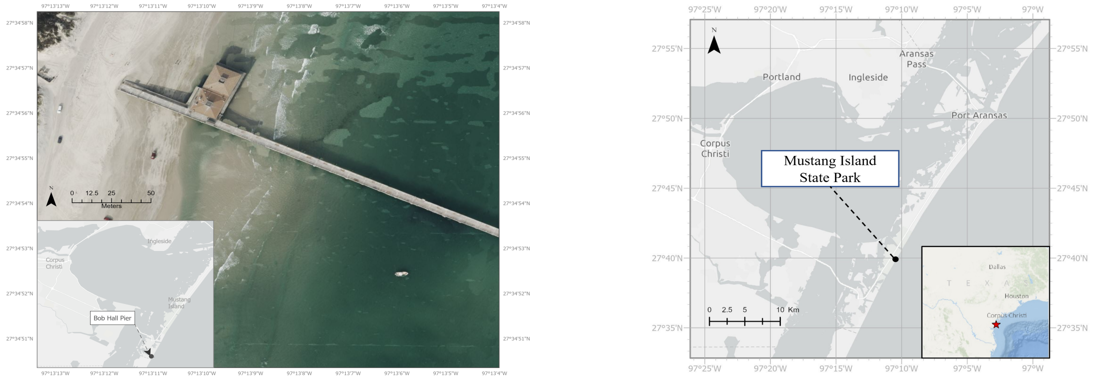
Additionally, conventional SfM-MVS computations produce a dense point cloud of the entire study area and its objects. Depending on the surveyed area's extent, this dense reconstruction can require significant computation time and resources [2], [3]. However, in specific applications, only certain objects or surfaces, such as buildings, roads, or bare ground, may be of interest for in-depth analysis. Processing image pixels outside these regions of interest adds a substantial computational burden, reducing the efficiency of the SfM-MVS workflow. This issue is especially significant when high temporal resolution is needed for reconstruction and mapping.

Moreover, existing SfM-MVS software suites lack robust and effective built-in tools for essential point cloud analysis tasks, such as terrain points filtering, point cloud segmentation, and object extraction. Although commercial software like Agisoft Metashape (Agisoft LLC, Russia) and Pix4Dmapper (Pix4D SA, Switzerland) have developed techniques and tools for filtering and classifying dense point clouds, their results are often unsatisfactory [3], [4]. This deficiency undermines the full potential of the SfM-MVS workflow for efficient reconstruction and mapping. As a result, using commercial or open-source point cloud processing software becomes essential for various analysis tasks. However, this approach presents challenges, including the need for capital investment in software and hardware, additional processing time, and evaluating the efficiency of the point cloud processing algorithms [5].

This study employs the SfM-MVS photogrammetry workflow previously proposed by the same authors in [3] and [6], with a focus on efficiently mapping two study sites. The workflow suggests utilizing semantically segmented images to guide the SfM-MVS processing of overlapping images, enabling simultaneous scene reconstruction and point cloud segmentation. The objective is to enhance the SfM-MVS processing workflow, improving reconstruction and mapping efficiency by increasing computational effectiveness and automation in generating geospatial products.

The presented approach utilizes pre-existing high-

M. Pashaei, M. J. Starek, J. Berryhill, and J. Pilartes-Congo are with the Conrad Blucher Institute for Surveying and Science, Texas A&M University-Corpus Christi, Corpus Christi, TX, 78412 USA (e-mail: mohammad.pashaei, michael.starek, jacob.berryhill@tamucc.edu, jcongo@islander.tamucc.edu).



a. Bob Hall Pier.

b. MISP.

Fig. 1: Bob Hall Pier (a) and MISP (b) study sites.

performance convolutional neural network (CNN) models, developed within deep learning (DL) frameworks, for pixel-wise image labeling, commonly referred to as semantic image segmentation [7]. The resulting segmented images are then integrated into the SfM-MVS processing workflow as binary image masks. These masks precisely delineate a specific set of pixels, subsequently influencing the input data for the SfM-MVS processing steps. Depending on the application, a CNN model may predict a set of image masks to exclude pixels associated with moving objects from entering the SfM-MVS processing. Additionally, another set of image masks can be predicted to enable dense reconstruction of specific targets in the study area, such as buildings, roads, or bare ground.

The proposed workflow tackles two challenging tasks. The first task involves reconstructing a narrow pier located over the dynamic surf zone within coastal water, where pixels associated with water are automatically excluded from processing to ensure a robust SfM-MVS solution. The second task entails simultaneous scene reconstruction and segmentation.

II. Materials and method

A. Materials

High-resolution UAS imagery was collected over two study sites, as depicted in Fig. 1, for image-based 3D reconstruction and mapping using SfM-MVS photogrammetry. The first site is Bob Hall Pier, a 1,240-foot elevated structure extending over the Gulf of Mexico on North Padre Island, Texas, USA. It was surveyed shortly after Hurricane Hanna in 2020, aiming to create a 3D model of the pier post-hurricane. A DJI Phantom 4 multi-rotor UAS, equipped with a 20-megapixel RGB camera, was deployed to capture 1570 aerial images over the site, achieving a ground sampling distance (GSD) of 1 cm using real-time kinematic (RTK) global navigtation satellite system (GNSS) mode for image geotagging.

The second study site is situated along a stretch of Gulf-facing sandy beach located in Mustang Island State Park (MISP), Texas, USA. This site requires a classified (segmented) dense point cloud covering various targets, including natural and built structures, as well as different surfaces. To achieve this, a Vertical Take-Off and Landing (VTOL) Wingtra One UAS equipped with a 42-megapixel RGB camera was deployed to capture $2,343$ aerial images with a ground sampling distance (GSD) of 1.5 cm over a 1.5 square kilometer area. The main targets in this study site include moving objects (water, people, vehicles on the beach), bare ground, vegetation, and man-made structures such as wooden walkways and pavilions located adjacent to the shoreline. The objective is to exclude moving objects from the SfM-MVS processing while simultaneously generating a dense, classified (segmented) point cloud through MVS computations for the remaining targets.

B. Method

Fig. 2 outlines the proposed processing workflow for image-based reconstruction and mapping using SfM-MVS photogrammetry. Initially, a CNN model, such as the UNet architecture, designed for semantic image segmentation, is employed and fine-tuned to predict semantic labels for each pixel within the acquired UAS image set, pertaining to the required target categories for reconstruction and mapping. The fine-tuning of model parameters (weights) is carried out through transfer learning, using a small set of training data sampled from the collected UAS image set. Subsequently, binary masks are generated from the resulting segmented images, specifying the subset of pixels in the overlapping image set that will be utilized in the subsequent SfM-MVS computations.

In this experiment, 100 images are selected from the initial UAS image set to fine-tune a pre-trained UNet model through transfer learning. The model's performance is evaluated using 20 images and their corresponding ground truths. The primary objective is binary image segmentation, specifically classifying image pixels into water and non-water categories. Subsequently, the original UAS image set, along with the corresponding binary masks created from the binary image segmentation, is used to exclude water pixels from the SfM-MVS processing. Following this, the same UNet model undergoes fine-tuning for multi-class classification using 300 images chosen from the second UAS image set, with 100 images reserved for evaluating the model's performance. Binary masks, pertaining to the required

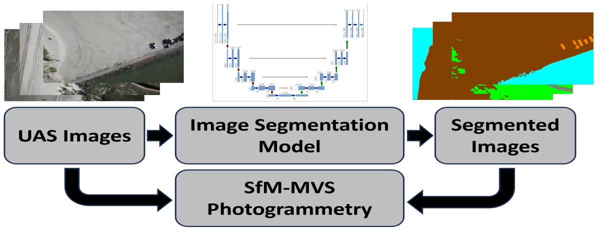


Fig. 2: Proposed technique for the UAS-SfM mapping.

TABLE I: Image segmentation error.

Study Site 1: binary image segmentation				
Class	**Water**	**Non-water**	**-**	**-**
Error (%)	2.1	1.0	–	–
Study Site 2: multi-class image segmentation				
Class	**moving-object**	**vegetation**	**ground**	**man-made**
Error (%)	1.50	6.40	5.85	8.20

target class(es) for reconstruction, are prepared by utilizing the multi-class segmented images. These masks are then employed in selective dense reconstruction through the MVS processing.

Commercial SfM-MVS photogrammetry software, such as Agisoft Metashape and Pix4Dmapper, enables the application of manually generated binary image masks onto input UAS images. This feature allows for excluding specific portions (image pixels) of the surveyed area from SfM processing and dense reconstruction. The proposed approach utilizes this functionality within the Agisoft Metashape software to improve efficiency and automation in UAS-SfM photogrammetry mapping.

III. Results and Discussion

This section presents quantitative and qualitative results from the proposed approach in this study. Table I offers an overview of the mislabeled pixel rates for each class category in both binary and multi-class image segmentation tasks applied to the UAS image sets from the first and second study sites. Regarding the misclassification rates for various target classes outlined in the table, the transfer learning of the UNet model using selected UAS images has consistently shown high performance in pixel-wise labeling across image sets for both study sites. It is noteworthy that the relatively higher error for vegetation and ground can be primarily attributed to the indistinct borders between these targets and the classifier's confusion over sparsely vegetated surfaces. Additionally, the elevated error in classification of built structures mainly results from the scarcity of such structures in the surveyed study area. Fig. 3 illustrates the quality of the segmented images predicted in the UNet model for some test images, where the middle row displays the true labels for pixels in each UAS image.

The classifier also shows lower performance in detecting pixels belonging to other moving objects in that class, such as cars and people in the surveyed scene, which is reasonable due to the scarcity of those objects in the UAS images.

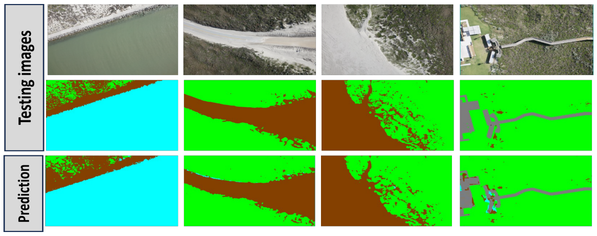


Fig. 3: An illustration of the semantic image segmentation result for sample test images.

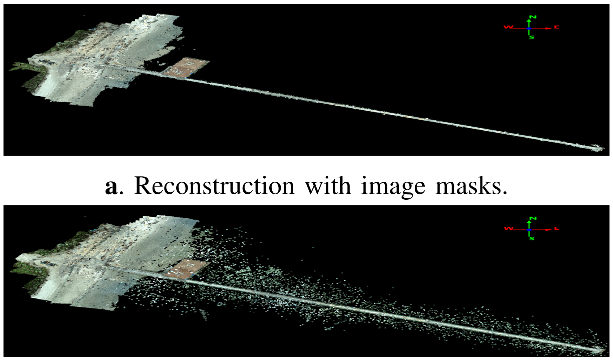
It's worth noting that the effectiveness of supervised classification models is influenced by numerous factors. Key determinants include the quality of the ground truth labeling used for training, data representation, initialization of model parameters and hyperparameters, and the choice of the model optimizer. Using more advanced CNN architectures for semantic image segmentation could enhance performance on the same dataset. However, concerning Table I, the relatively low error rate in segmenting the surveyed area serves as a reliable metric that offers a rough approximation of the quality of characterization (segmentation) of the study area within the image space, as well as subsequent image-based reconstruction and mapping through the SfM-MVS photogrammetry processing.

Table II presents key metrics for the SfM-MVS solution of the pier reconstruction over the dynamic coastal water at the first study site, including the total number of valid tie points, the total number of projections of valid tie points, the root mean square (RMS) of reprojection error averaged across all tie points on all images, and the total number of points in the generated dense point cloud. The table provides a comparative analysis between solutions with and without the utilization of binary image masks of the water pixels in the processing workflow. Fig. 4 illustrates the dense reconstruction of the pier in the study area for both reconstruction scenarios. As indicated in the table, incorporating image masks in SfM processing led to an approximately 50% reduction in the RMS of the projection, which is a crucial metric quantifying the quality of the SfM least-squares bundle adjustment (BA) computations for the UAS image alignment and the accuracy of estimated camera parameters. Notably, the total number of valid tie points and the total number of projections of valid tie points exhibited an increase when pixels associated with water in the UAS images were excluded from the SfM processing. Referring to Table II and Fig. 4a, approximately nine million points in the dense point cloud in the SfM-MVS process without image masks indicate noisy points surrounding the pier. This noise is attributed to the high uncertainty in the matching process, primarily influenced by the presence of moving water in the surveyed scene.

Employing binary masks specifically designed to filter water into the SfM-MVS workflow for the second study site results in precise mapping of the shoreline, as illustrated in Fig. 5a. The incorporation of these masks effectively removes water from the SfM processing, enabling accurate 3D mapping of the shoreline. This accuracy is evident in both the quality

TABLE II: SfM-MVS solution for Pier reconstruction.

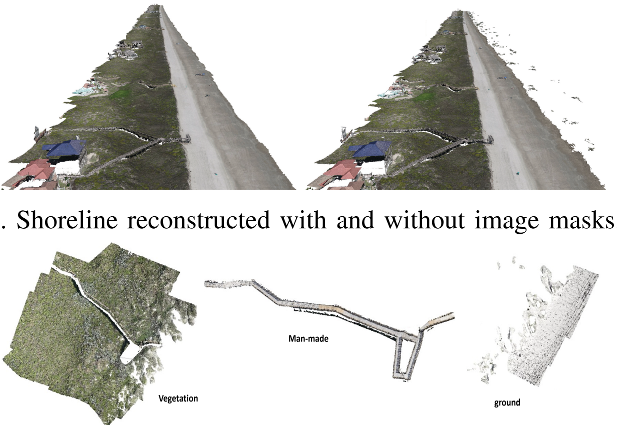
SfM-MVS solution without image masks			
Tie points	**Projections**	**Reprojection error**	**Dense point cloud**
$563.5K$	$1.9M$	$1.1\ pix.$	$79.2M$
SfM-MVS solution with image masks			
Tie points	**Projections**	**Reprojection error**	**Dense point cloud**
$663,5K$	$2.5M$	$0.65\ pix.$	$70.3M$



a. Reconstruction with image masks.

b. Reconstruction without image masks.

Fig. 4: Pier reconstructed with and without image masks.



a. Shoreline reconstructed with and without image masks.

b. Segmented dense point cloud.

Fig. 5: Visualization of the reconstructed shoreline and segmented point cloud.

assessment of detected water-ground boundaries (shoreline location) in the ground truth images and the segmentation error rates provided in Table I.

Furthermore, Fig. 5b provides a visualization of the segmented dense point cloud for a small portion of the surveyed area, generated simultaneously with the dense reconstruction of the required targets in the study site, namely bare ground, vegetation, and built structures, utilizing sets of corresponding binary image masks. This illustrates the effectiveness of employing image masks to facilitate the generation of segmented scenes alongside dense reconstruction directly through the MVS processing as opposed to after dense point cloud generation as is typically done.

It's crucial to highlight that employing advanced image segmentation models can greatly enhance performance, especially in delineating surface and object boundaries of various scales. These models excel in accurately labeling pixels on captured surfaces and objects appearing at different scales in the collected aerial images. For instance, they can identify sets of pixels within the collected UAS image set with a certain GSD that represents exposed ground surfaces of varying sizes within vegetated areas. Efficiently extracting terrain points within these areas and beneath the vegetation canopy poses a challenging yet crucial task for producing precise digital terrain models (DTMs).

IV. Conclusion

This study presents an approach aimed at enhancing computational efficiency and automation in reconstruction and UAS mapping tasks using SfM-MVS photogrammetry. In this proposed workflow, pixels are selectively included in the SfM-MVS computation based on their predicted labels from the segmentation task, enabling scene reconstruction and focused dense mapping. Further research is needed to thoroughly assess the effectiveness of this approach in mapping and characterizing surveyed environments with varying complexities in aerial mapping.

V. Acknowledgements

This work was supported in part by the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce under Award NA18NOS4000198, and in part by the National Science Foundation (NSF) under Award 2112631 and 2150351. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding sources.

References

[1] Matthew J Westoby, James Brasington, Niel F Glasser, Michael J Hambrey, and Jennifer M Reynolds. 'structure-from-motion'photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179:300–314, 2012.

[2] EF Berra and MV Peppa. Advances and challenges of uav sfm mvs photogrammetry and remote sensing: Short review. In *2020 ieee latin american grss & isprs remote sensing conference (lagirs)*, pages 533–538. IEEE, 2020.

[3] Mohammad Pashaei, Michael J Starek, and Jacob Berryhill. Application of semantic image segmentation for efficient uas-sfm photogrammetry mapping. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 6983–6986. IEEE, 2023.

[4] Carlos Becker, Nicolai Häni, Elena Rosinskaya, Emmanuel d'Angelo, and Christoph Strecha. Classification of aerial photogrammetric 3d point clouds. *arXiv preprint arXiv:1705.08374*, 2017.

[5] Mustafa Zeybek and İsmail Şanlıoğlu. Point cloud filtering on uav based point cloud. *Measurement*, 133:99–111, 2019.

[6] Natthapol Saovana, Nobuyoshi Yabuki, and Tomohiro Fukuda. Automated point cloud classification using an image-based instance segmentation for structure from motion. *Automation in Construction*, 129:103804, 2021.

[7] Xiaohui Yuan, Jianfang Shi, and Lichuan Gu. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169:114417, 2021.