3D Reconstruction with Drone Images: Optimization by Reinforcement Learning

Thiago João Miranda Baldivieso
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: thiagojmb@ime.eb.br

Taise Grazielle da Silva Batista
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: taisegbs@ime.eb.br

Luiz Carlos Pacheco Rodrigues Velho
Department of Computer Graphics
Institute for Pure and Applied Mathematics
Rio de Janeiro, Brazil
email: lvelho@impa.br

Paulo Fernando Ferreira Rosa
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: rpaulo@ime.eb.br

***Abstract*—This paper aims to develop three-dimensional reconstructions using aerial images in different environments and using dedicated software. The subject is relevant to this conference because, due to the characteristics of photogrammetry with UAV/drone, they provide easy access, accuracy, and time-saving and mission equipment for civil and military applications. During the study, experiments were carried out with aircraft in an external and internal environment. After acquiring aerial images, the reconstruction was carried out in specific photogrammetry software, with typical commercial and open-source software, followed by a qualitative evaluation of the results. Concluding with indications for improvements and future research work related to artificial intelligence techniques using machine learning and reinforcement learning to optimize.**

Keywords*—*Dimensional Reconstruction; UAV; Aerophotogrammetry

I. INTRODUCTION

With increasing demand and actual needs, the functions and performance of Unmanned Aerial Vehicles (UAVs) are continually advancing. Technological advances mainly drive the area of microprocessors, sensing, communications, and open demands in the areas of computer vision and computer graphics in the reconstruction of objects and threedimensional environments. Furthermore, applications with autonomous and semi-autonomous UAVs, characterized with total or partial independence from human operators, provide greater visibility in the image, as it is not necessary for the operator to aim the aircraft during the entire mission. Furthermore, in the scope of applications in the civil and military sectors, it has reduced operating costs and encouraged financing initiatives in the area.

The use of unmanned aerial vehicles for three-dimensional mapping and reconstruction requires mission planning considering the object or environment from which the images will be taken and factors that can influence the process, such as weather, lighting, target geometry, camera calibration, and type of aircraft used. With the studies carried out based on reference works, it is possible to observe that some parameters are not considered, such as attitude control, calling attention to observe the parameters that influence and if not using any parameter is valid for the process as a whole.

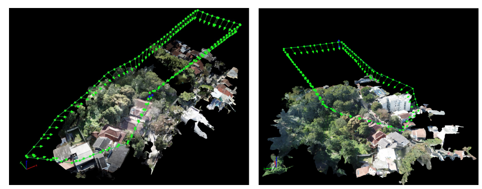

Figure 1. Mapped and reconstructed outdoor area.

Three-dimensional reconstruction is a highly researched area in computer vision and scientific visualization. Its objective is to obtain a three-dimensional geometric representation of environments or objects, making it possible to inspect details, measure properties, and reproduce them in different materials. Applications with UAVs can help in architecture, 3D cartography, robotics, augmented reality, conservation of monuments, and historical heritage [2]. There are several ways to get information related to the 3D geometry of an object, environment, or body. They can be acquired by laser scanning, photographs, sonar, tomography, and 3D sonar. On the other hand, photo-based systems make 3D reconstructions from a single photo or with several photos at different angles, using multiple photos, which after image registration, consists of transforming different sets of data into a coordinate system. After this step, visual reference points are defined, automatically generated by the reconstruction software, or entered manually. To establish typical visual landmarks in the scene to identify joint edges of the object to be processed in the photographs. From the processing of this information, three-dimensional geometry is obtained. In addition, each photograph is registered by the UAV with information about the location of the Global Positioning System (GPS) sensor and the time of capture, information that is also considered in the processing to obtain the model's georeferencing. Figure 1

 ISBN: 978-1-61208-918-8

presents a result of the 3D reconstruction of an external scene. The green dots are the poses of the drone's camera.

Given the above, the project in which this work is part aims to relate images captured by UAVs to develop three-dimensional reconstructions from them and the continued study for optimization with learning by reinforcement.

Section 2 of this work presents the methodology with some relevant concepts for understanding. Then, in section 3, some experiments that were carried out are described. Then, in section 4, the results were discussed, and finally, in section 5, the conclusion with the analysis of what was proposed.

II. METHODOLOGY

In recent years, aircraft used for terrain mapping in civil and military applications have been widely explored, especially unmanned aircraft and their use related to three-dimensional reconstruction.

Computer vision is defined as the science and technology of machines that see. In [3], the author develops theory and technology for the construction of artificial systems for obtaining information from images or any multidimensional data

The concepts of machine vision were initially restricted to the construction of lenses and cameras for image capture and operations. However, over the past few years, this reality has been modernized due to the growth of artificial intelligence and the application of the concept of neural networks, along with the improvement of studies on the self-progression of algorithms, known as machine learning [4]. Soon then, computer vision can be included in a sub-area of Artificial Intelligence that addresses how machines see the environment. Furthermore, a body of knowledge that seeks the artificial modeling of vision can also be defined to replicate its functions through advanced software and hardware development.

Applied 3D reconstruction software uses the Structure for Motion method [5], which uses said relative motion for the inference about the 3D geometry of the object to be reconstructed. The methodology also encompasses bundle adjustment, which initially compares the key point descriptors identified in the images to determine between two or more similar images. Then a procedure optimization is performed to infer the camera positions for the collection of images.

Structure From Motion (SfM) is a range imaging technique studied in machine vision and visual perception. The SfM methodology uses this relative motion to infer the 3D geometry of the object to be reconstructed. It takes into account the point trajectories of the object in the image plane and allows the determination of the 3D shape and movement that best reproduces most of the estimated trajectories. The process is similar to stereoscopic vision in that it is done to obtain two or more images of a scene from different points of view [6].

Consider a picture arrangement comprising of K pictures I_k, with K = 1,...,K. Leave A_k alone the 3 x 4 camera framework comparing to picture I_k. Utilizing the compared highlight focuses, the boundaries of a camera model A_k are assessed for each casing [5]. As displayed in Figure 2, for each element track a relating 3D item point not really set in stone, bringing about set of J 3D article focuses P_j, with j = 1,...,J, where:

$$p_{j,k} \simeq A_k Pj \tag{1}$$

Accordingly, the 2D component focuses $p_{j,k} = (p_x, p_y, 1)^T$ and 3D item focuses $P_j = (p_x, p_y, p_z, 1)^T$ are given inhomogeneous directions.

The camera network A can be factorized into

$$A = KR\,[I| - C] \tag{2}$$

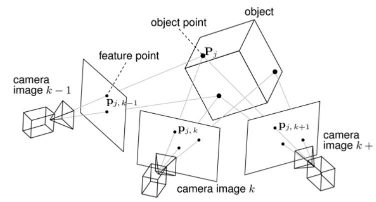

Figure 2. Result after structure-from-motion estimation. The projection of a 3D object point Pj in the camera image at time k gives the tracked 2D feature point $P_{j,k}$ [6]

The 3 x 3 adjustment lattice K contains the inherent camera boundaries (e.g., central length or chief point offset), R is the 3 x 3 turn framework addressing the camera directly in the scene, and the camera place C depicts the situation of the camera in the scene. SFM is considered state of the art in reconstruction software because it solves camera poses and lens calibration in addition to defining geometries [5].

3D reconstruction is an old problem. Ways to improve the process have become a focus of research with reconstruction forms using current AI that reinforces the use of SFM by programs. In [4] provides a comprehensive survey of recent developments in this field, works that use deep learning techniques to estimate the 3D shape of generic objects. The work provides an analysis and comparison of the performance of some critical documents, summarizing some of the open problems in this field and discussing promising directions for future research.

III. EXPERIMENTS

We performed several experiments, which can be seen in Table I, outdoor and indoor experiments, using UAVs and also ground cameras. After the images were taken, reconstructions were carried out in different software such as PIX4D, Metashape, OpenDroneMap, and Colmap using default configurations.

BOC 60 is the new campus of the Institute of Pure and Applied Mathematics (IMPA) to be built in Jardim Botanico, in the ˆ south of the city of Rio de Janeiro.

In the flight test of Figure 3, it was with the help of GPS in an urban setting with different types of buildings, vegetation, and complex shapes, resulting in three-dimensional models with high processing demand when performing the image matching step. These outdoor flights were carried out in partnership with IMPA with the UAV research group of

 ISBN: 978-1-61208-918-8

TABLE I
GENERATED DATASETS, IMAGES AND RECONSTRUCTION SOFTWARES USED.

Dataset generated	Acquisition Device	Images	PIX4D	Meta-shape	ODM	RC
Crystal's valley	Mavic Pro	337	x	x	x	
BOC 60 - High Res.	Mavic Pro	302	x	x	x	
BOC 60 - Med. Res.	Mavic Pro	169	x	x	x	
BOC 60 - Low Res.	Mavic Pro	138	x	x	x	
LARC	Sub-250	150	x	x		
PIRF - Fan Scene	Tello	62				x
PIRF - Human Scene	Tello	50			x	x
PIRF - Bags Scene	Smartphone	217	x			
Object - Plant	Tello	35	x	x		x
Object - Robot	Smartphone	154	x			
Object - Castell	Smartphone	64	x	x	x	x

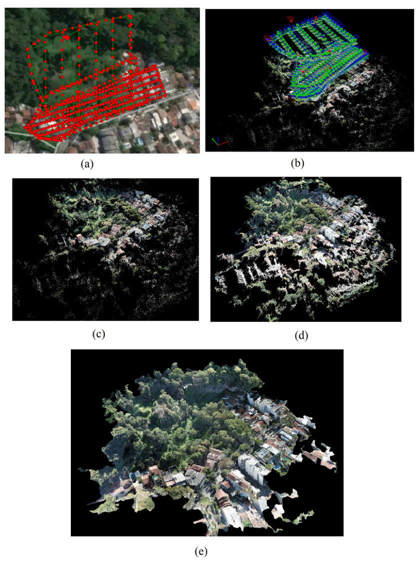

Figure 3. BOC 60 Steps to Rebuild PIX4D software; (a) Snapshot points on the map; (b) 3D image taking points; (c) Tie Points; (d) Dense cloud of points; (e) Textured 3D Model.

Figure 4. Visual comparison of three-dimensional reconstruction results.

the Laboratory of Robotics and Computational Intelligence of IME to obtain images aimed at aero photogrammetry and create a dataset. More information about it is in [7].

The outdoor experiment was divided into three missions aiming to obtain a differentiated resolution, high resolution, medium resolution, and low resolution, with the variation of height and number of photos obtained.

Next, in Figure 3, the steps for reconstruction are observed, in (a) all points of image collection by the UAV are gathered, then in (b) the camera pose is presented (position + orientation) in a three-dimensional plane; (c) shows the initial step in which the tie points are characteristic points mapped between the images; (d) are the initial points gathered clustered with neighboring points resulting in a dense cloud of points; and finally in (e) the three-dimensional object is obtained in which a mesh structure and texture connect the cloud of points is applied based on a montage of images, forming an object close to urban reality.

With the creation of the dataset of images, the next step was to reconstruct the mapped terrain. For this, three different software were used for the three-dimensional reconstruction. They are PIX4D (whose process was discussed earlier), Metashape, and OpenDroneMap. The visual comparison between results can be seen in Figure 4. Another outdoor experiment can be visualized in our generated dataset [7].

Another experiment was related to the use of a Ground Truth model of a medieval castle. First, the object was printed to obtain the physical object. The model is available from the website Thingiverse, modeled on two castles: Schloss Lichtenstein and Neuschwanstein Castle, both located in Germany. In addition, authors of the design and graphic modeling of the object make it available for download through [8]. Next, in Figure 5, you can see the file to be printed and the part that has already been printed. The model is approximately one meter high and is divided into 22 pieces, the model used in the experiment was glued by the assembly instructions, but due to deformations caused during printing, some parts need finishing and painting to be as close as possible the virtual model.

Figure 5. Medieval Castle experiment, 3D file visualization, printed parts and image of dataset.

The experiment with the medieval castle printed in 3D was carried out in order to generate a reconstructed model to later compare with a new experiment to be carried out using a UAV on an indoor scene. In this partial experiment, the acquisition of the image collection was performed using a smartphone camera. Soon after obtaining the images, processing in three-dimensional reconstruction software was performed. With these results and the following experiments, the aim is to apply reinforcement machine learning algorithms to optimize the generated three-dimensional models.

 ISBN: 978-1-61208-918-8

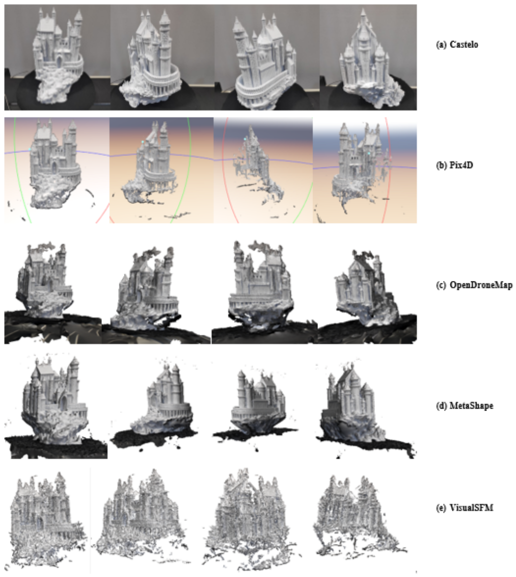

Figure 6. Medieval Castle experiment, reconstruction using different tools.

IV. Discussion

With the particularities of each tool, the clouds of generated points present their differences, and it becomes interesting to compare them for analysis of the results.

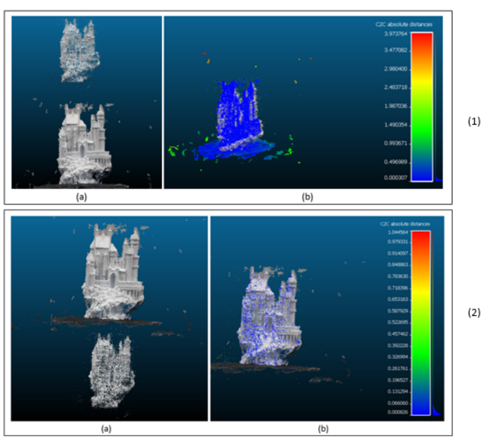

Figure 7. Comparison point cloud Metashape and OpenDroneMap. (1) Metashape reference. (2) OpenDroneMap reference. (a) Insertpoint cloud; (b) Generated heat map.

CloudCompare software was used for comparison between point clouds [9]. CloudCompare is a point cloud processing tool with multiple metrics; it is an open-source and free project with a framework that provides a set of essential tools for manually editing and rendering 3D point clouds and triangular meshes.

The initial analyses were carried out from the reconstruction of the image set of the medieval castle that obtained a good result. Since not all tools make the files available to be exported, the comparison was performed with the files generated by the Metashape and OpenDroneMap tools.

For the research, an analysis of the distance between points with heat cloud generation was performed, the clouds generated in each tool were inserted, and the analysis was performed in two stages. The first step was using the cloud generated by Metashape as a reference cloud. The second step was using the cloud generated by OpenDroneMap as a reference cloud. The result obtained is shown in Figure 7.

It is observed that the distance between the points of the clouds presents a significant difference, and the scale and orientation factors of each cloud must be treated with due care in the comparison. With the result generated, it is possible to qualitatively analyze the generated clouds and identify the software that presents better performance and the need for improvements through machine learning by reinforcement.

V. Conclusion and Future Work

The contribution made by this project includes the creation of datasets with scenes and 3D objects obtained through reconstruction and images captured by drones. These data are available to the academic community and have several capture devices, processed by exposed dedicated software in Table 1. In the continuation of the work, it is expected to use these data for optimization experiments with machine learning and reinforcement learning to improve the distortions caused during image processing and increase the visible accuracy of the three-dimensional models.

Acknowledgement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work was carried out with the support of the Cooperation Program Academic in National Defense (PROCAD-DEFESA). It is also aligned with the cooperation project between BRICS institutions related to computer vision and applications of AI techniques.

References

[1] E. Casella, et al. "Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques". Coral Reefs, vol. 36, pp.269–275, 2017.

[2] E. Colica, et al. "Using unmanned aerial vehicle photogrammetry for digital geological surveys: case study of Selmun promontory, northern of Malta". Environ Earth Sci 80, pp. 551, 2021.

[3] I. Craig, Vision as process. Robotica, Cambridge University Press, vol. 13, n. 5, pp. 540, 1995.

[4] X. Han, H. Laga and M. Bennamoun, "Image-Based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 05, pp. 1578-1604, 2021.

[5] S. M. Seitz, B. Curless, and J. Diebel, and D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006.

[6] C. Kurz, T. Thormählen, and H. Siedel, "Visual Fixation for 3D Video Stabilization". Journal of Virtual Reality and Broadcasting, pp. 12, 2011.

[7] L. C. P. Velho, Drone Datasets. 2020. Available at: https://www.visgraf.impa.br/dds/boc60/index.html [Retrieved: November, 2021]

[8] Boldmachines, Medieval Castle, 2018. Available at: https://www.thingiverse.com/thing:862724 [Retrieved: November, 2021]

[9] CloudCompare - Open Source Project. Available at: https://www.cloudcompare.org [Retrieved: November, 2021]

 ISBN: 978-1-61208-918-8