3D Reconstruction with Drone Images: optimization by reinforcement learning

1st Thiago João Miranda Baldivieso
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: thiagojmb@ime.eb.br

2nd Taise Grazielle da Silva Batista
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: taisegbs@ime.eb.br

3rd Luiz Carlos Pacheco Rodrigues Velho
Department of Computer Graphics
Institute for Pure and Applied Mathematics
Rio de Janeiro, Brazil
email: lvelho@impa.br

4th Paulo Fernando Ferreira Rosa
Department of Defense Engineering
Military Institute of Engineering
Rio de Janeiro, Brazil
e-mail: rpaulo@ime.eb.br

***Abstract*—**

Studies related to computer vision involving three-dimensional reconstruction are currentlyincreasing and demonstrate the importance of the area that is a subdomain of photogrammetry. Applications with UAVs (Unmanned Aerial Vehicles) for terrain mapping using photogrammetry to create three-dimensional environments facilitate structural analyzes and make it possible to obtain topographic information from captured land surfaces,using modern methodologies applicable in both civil and military areas. The subject hasrelevance, because due to the characteristics of photogrammetry with UAVs, they provideeasy access, precision and savings in mission time and in equipment. This dissertation aimsto develop three-dimensional reconstruction using aerial images in different environments.During the study, experiments were carried out with aircraft in external and internal environments, after the acquisition of aerial images, reconstruction was performed using specific photogrammetry software, with caracteristic commercial and open-source softwares, followed by a qualitative evaluation of the results. Concluded with indications of improvements and continued work for research related to artificial intelligence techniques using machine learning and reinforcement learning to optimization.

***Index Terms*—Three-dimensional reconstruction. UAV. Aerophotogrammetry.**

I. Introduction

With increasing demand and real needs, UAV functions and performance are continually progressing. Technological advances mainly boost the area of microprocessors, sensing, communications and also open demands in the areas of computer vision and computer graphics in the reconstruction of objects and three-dimensional environments. Applications with autonomous and semi-autonomous UAVs, characterized with total or partial independence from human operators, provide greater visibility in the image because it is unnecessary for the operator to aim at the aircraft during the entire mission. Within the scope of applications in civil and military sectors, it has enabled the reduction of operating costs and encourages the financing of initiatives in the area.

The use of UAVs for imaging aimed at mapping and three-dimensional reconstruction requires planning the mission according to the target object or environment, as well as some factors that must be taken into account during the mission, such as weather, lighting, target geometry , mission trajectory, camera calibration and type of aircraft to be used. Among studies carried out using UAVs for three-dimensional reconstruction, it was observed that works used as reference do not take into account some parameters of the aircraft, such as attitude control during trajectory planning carried out in the imaging mission.

Three-dimensional reconstruction is a highly researched area in both 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 [7], measure properties and reproduce them in different materials . Applications with UAVs can help in the areas of architecture, 3D cartography, robotics, augmented reality, conservation [6] of monuments and historical heritage. Information related to the 3D geometry of an object, environment or body can be acquired by laser scanning, photographs, sonar, tomography and 3D sonar. The laser scanning system measures the distance from the source based on the delay between the emission of the light signal and its return. While photo-based systems do 3D reconstruction from one photo (in which the face of the object is determined and then extruded) or with multiple photos at different angles. When using multiple photos, in which after image registration, it consists of the process of transforming different sets of data into a coordinate system. After this step are defined points of visual references, automatically generated by the reconstruction software or inserted manually, establishing common visual landmarks in the scene, in order to identify common edges of the object to be processed in the photographs, from the processing of this information, three-dimensional geometry is obtained. In

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

addition, each photograph is registered by the UAV with information on the location of the GPS (Global Positioning System) sensor and the moment of capture, such information is also taken into account in the processing to obtain the georeferencing of the model.

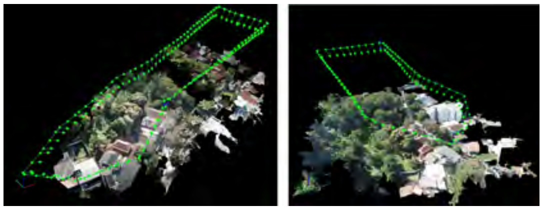

Fig. 1. Mapped and reconstructed outdoor area.

In view of the above, the project in which this work is inserted has as purpose: a related of the use of images captured by UAVs for the reconstruction of 3D scenes, aiming to develop three-dimensional reconstruction using UAV images and the continued study for optimization with reinforcement learning. In the course of this work, we will address its theoretical foundation, related scientific work, mission methodologies, experiments and expected progress.

II. Methodology

The use of aircraft for terrain mapping in civil and military applications has been widely explored over the past few years, especially unmanned aircraft and their use related to three-dimensional reconstruction.

This section presents theoretical references related to the theme raised in this article.

A. *Computer Vision and 3D Reconstruction*

Computer vision is defined by [1] as the science and technology of machines that see. She develops theory and technology for the construction of artificial systems for obtaining information from images or any multidimensional data.

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

This technology has been widely used in the military, but with confidential aspects. Some applications are known to the general public such as detecting enemy units and homing missiles. However, domestic applications such as facial recognition for security, autonomous vehicles and in the industrial field, the use of computer vision has been widespread and applied, bringing great benefits and technological and academic advances.

Applied 3D reconstruction software uses the SFM method (Structure for Motion) [8] 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 keypoint descriptors identified in the images for determine betwen two or more similar images. Then a procedure optimization is performed to infer the camera positions for the collection of images.

Structure From Motion is a range imaging technique studied in the fields of machine vision and visual perception. The SfM - Structure from Motion methodology uses the referred relative motion for the inference about the 3D geometry of the object to be reconstructed: considering trajectories of points of the object in the image plane, the SfM method enables the determination of the 3D shape and movement that best reproduces most of the estimated trajectories. Its process is similar to that of the stereo vision in that it is made to obtain two or more images of a scene from points from different views [8].

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 comparing highlight focuses, the boundaries of a camera model A_k are assessed for each casing [2]. 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 in homogeneous directions.

The camera network A can be factorized into

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

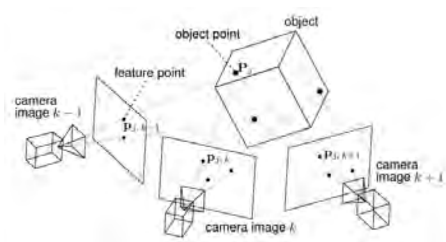

Fig. 2. Result after structure-from-motion estimation. The projection of a 3D object point Pj in the camera image at time k gives the tracked 2D feature point $P_{j,k}$ [2]

Where 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 direction in the scene, and the camera place C depicts the situation of the camera in the scene.

III. Experiments

We performed several experiments, which can be seen in Figure , 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. BOC 60 is the new campus of IMPA (Institute of Pure and Applied Mathematics) to be built in Jardim Botânico, in the south of Rio de Janeiro. The land has 251,824.72 m^2 and the construction area will have 8,140.30 m2 and includes auditoriums, offices for researchers and students, a library, classrooms and dormitories.

In this flight test, 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.

Outdoor flights were carried out in partnership with IMPA (Institute of Pure and Applied Mathematics) with the UAV research group of the Laboratory of Robotics and Computational Intelligence, to obtain images aimed at aerophotogrammetry and creation of *dataset*. More information about it in [3].

The experiment was divided into three missions aiming to obtain differentiated resolution, high resolution, medium resolution and low resolution, having as parameters 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 the cloud of points are connected by a mesh structure and texture is applied based on a montage of images, forming an object close to urban reality.

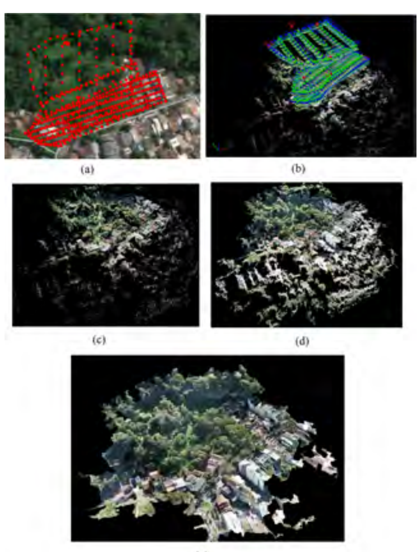

Fig. 3. BOC 60 Steps to Rebuild PIX4D software; (a) Snapshot points on the map; (b) 3D image taking points; (c) Tie Points; (d) Dense cloud of points; (e) Textured 3D Model.

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 6.

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

Another experiment to related using a model *Ground Truth* of a medieval castle, and printing the object 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. Authors of the design and graphic modeling of the object make it available for download through [4]. 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 in accordance with the assembly instructions, but due to deformations caused during printing some parts need finishing and painting to be as close as possible. to the virtual model.

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

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

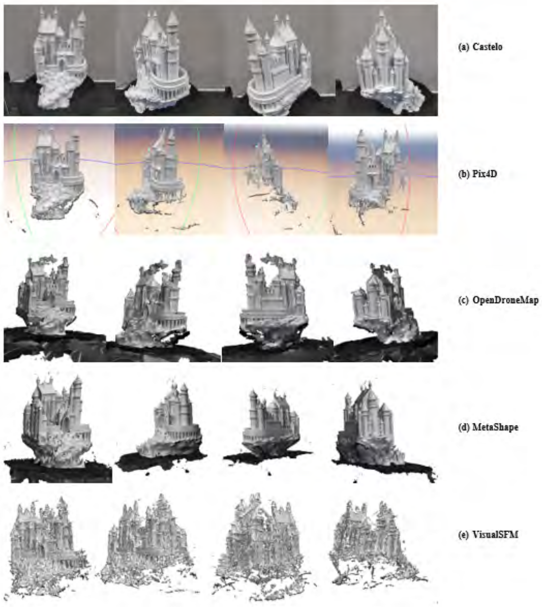

Fig. 6. Medieval Castle experiment, reconstruction using different tools.

CloudCompare software was used for comparison between point clouds [5]. 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 basic 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 and resulted in the presented in Figure 9.

The second step was using the cloud generated by OpenDroneMap as a reference cloud and resulted in the presented in Figure 8.

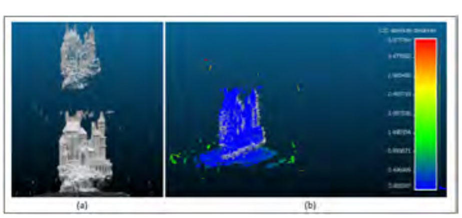

Fig. 7. Comparison point cloud Metashape and OpenDroneMap, Metashape reference. (a) Insertpoint cloud, Metashape cloud top, and OpenDroneMap cloud bottom; (b) Generated heat map.

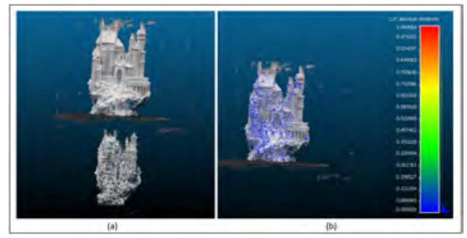

Fig. 8. Comparison of metashape and opendronemap points, OpenDroneMap reference. (a) Insertpoint cloud, OpenDroneMap cloud top, and metashape cloud bottom; (b) Generated heat map.

IV. Conclusion

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 has 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 in order to improve the distortions caused during image processing and also increasing the visible accuracy of the three-dimensional models.

Experimento	Dispositivo de aquisição	Imagens	PIX4D	Meta-shape	ODM	RC
Vale dos Cristais	Mavic Pro	337	x	x	x	
BOC 60 - Alta Res.	Mavic Pro	302	x	x	x	
BOC 60 - Méd. Res.	Mavic Pro	169	x	x	x	
BOC 60 - Baixa Res.	Mavic Pro	138	x	x	x	
LARC	Sub-250 Prot.	150	x	x		
PIRF - Cena Ventilador	Tello	62				x
PIRF - Cena humano	Tello	50			x	x
PIRF - Cena Malas	Celular	217	x			
Objeto - planta *outdoor*	Tello	35	x	x		x
Objeto - Robô Pioneer	Celular	154	x			
Objeto - Castelo	Celular	64	x	x	x	x

Fig. 9. Table with generated datasets, images and software used

References

[1] CRAIG, I. Vision as process.Robotica, Cambridge University Press, v. 13, n. 5, p. 540,1995.

[2] Kurz C., Thormählen T., Siedel H. 2011. Visual Fixation for 3D Video Stabilization. Journal of Virtual Reality and Broadcasting, 8(2):12 p.

[3] VELHO, L. C. P.Drone Datasets. 2020. Access em: 20 of September of 2021. Available on: https://www.visgraf.impa.br/dds/boc60/index.html.

[4] BOLDMACHINES.Medieval Castle. 2018. Acesso em: 20 of September of 2021. Available on: https://www.thingiverse.com/thing:862724.

[5] CloudCompare - Open Source Project. Access em: 23 of September of 2021. Available on: https://www.cloudcompare.org/

[6] Colica, E., D'Amico, S., Iannucci, R. et al. Using unmanned aerial vehicle photogrammetry for digital geological surveys: case study of Selmun promontory, northern of Malta. Environ Earth Sci 80, 551 (2021).

[7] Casella E, Collin A, Harris D, Ferse S, Bejarano S, Parravicini V, Rovere A (2017) Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques. Coral Reefs 36(1):269–275

[8] S. M. Seitz, B. Curless, J. Diebel, 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