Reconstrucción en 3D de contextos arqueológicos utilizando segmentación de imagen y estructura desde el movimiento



AUTOR

LUIS ENRIQUE MELO BARRERA

ALEXANDER CERÓN CORREA

Artículo de investigación presentado como requisito para optar al título de:

INGENIERO EN MULTIMEDIA

Tutor:

ALEXANDER CERÓN CORREA

UNIVERSIDAD MILITAR NUEVA GRANADA

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA EN MULTIMEDIA

BOGOTÁ, 06 DICIEMBRE 2021

3D reconstruction of archaeological contexts using image segmentation and Structure from Motion

Luis Enrique Melo Barrera
Facultad de Ingeniería
Universidad Militar Nueva Granada
Bogotá, Colombia
Email: est.luise.melo@unimilitar.edu.co

Alexander Cerón Correa
Facultad de Ingeniería
Universidad Militar Nueva Granada
Bogotá, Colombia
Email: alexander.ceron@unimilitar.edu.co

***Abstract*—Into the large set of techniques to elaborate 3D reconstructions, the Structure from Motion algorithms stand out because of its low cost and versatility, being these algorithms a series of computational procedures to the automated generation of dense point clouds from a set of photographs. The goal of this paper is to propose a workflow for the 3D reconstruction of archaeological objects and contexts employing a low cost UAV (Unmanned aerial vehicle) for the data capture phase, searching to evaluate and problematize the need of use expensive and specific hardware. After this data capture stage the 3D reconstruction process begins. This workflow is composed by three great phases: The generation of binary masks to select regions of interest, the production of dense point clouds and its transformation into a triangular mesh. Finally, sketch out the challenge that means aspire to determine which is the best of the generated reconstructions under quantitative frames.**

***Index Terms*—Archaeological reconstruction, digital heritage, triangular mesh, point cloud, image segmentation, SfM.**

I. Introduction

DIGITAL archaeology has gained relevance within archeology for allowing the virtual reconstruction in three dimensions of archaeological contexts for conservation and study purposes. To achieve these three dimensions reconstructions many techniques are employed, depending on what the study is trying to reach. The most frequently used technologies in this field are photogrammetry techniques [1], three dimensional scanning [2] , 3D modeling [3] and even LiDar technologies approaches. Although there are many technologies as mentioned before, there is a technique which highlights because of its versatility and low cost. The structure from motion technique consists into a set of algorithms and processes for the point cloud generation based on overlapping a series of images of the same object, seen from many viewpoints. Inside the polygonal meshes generation process using Structure from Motion is possible to employ binary masks to delimit regions of interest in the data set, as is pretended to show next up, the use of this binary masks has determinant repercussions in the visual quality of the generated model. The issue is in terms of visual quality and detail because it is searching to problematize how the quality of a virtual reconstruction is determined.

Fig. 1 shows an overview of the workflow proposed in this paper for the full 3D reconstruction pipeline.

Actually the binary masks creation process inside SfM must be done manually (when the object which will be reconstructed is unique, e.g), even though it is not a problematic matter when the data set size is small, it tends to turn cumbersome in function of the archaeological context complexity and the data set size. This paper seeks to propose a low cost method to the generation of triangular meshes of archaeological objects and contexts, and also a semiautomatic way to develop the binary masks creation process as the region of interest exists around an unique object.

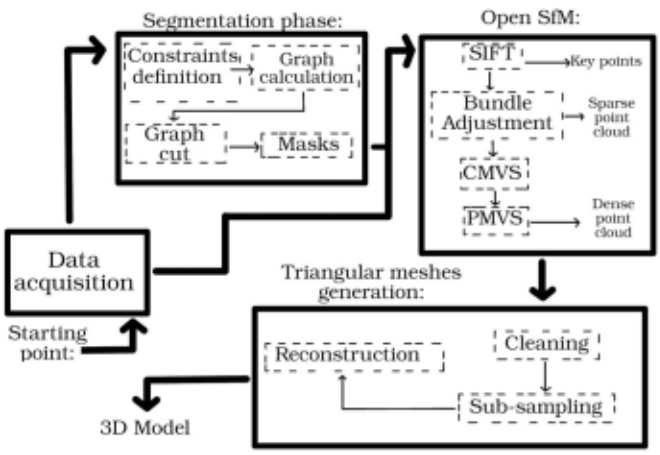


Fig. 1: Procedural framework for the generation of polygonal meshes

II. Materials and Methods

A. Hardware resources

A low cost UAV was used in this approach, the employed UAV model is SG906 Pro and was used for the data acquisition phase. This device has a 4k resolution camera, an 120° angle, a value of 85 mm as focal length and a two axis stabilizer. The computer where the reconstructions were made is an Acer Aspire 3 A315-34-C32Q, with Intel Celeron processor (2 cores and 2.6GHz of frequency), 4GB of RAM and 500 GB HDD.

978-1-6654-1668-9/21/$31.00 ©2021 IEEE

B. Software resources

Continuing with the above idea, all the software developed and employed in the present work is free and open source software (FOSS), inside the used software lies OpenSfM, OpenCV and MeshLab. Also the employed computer runs on GNU/Linux.

C. Image resources

The built data set was initially conformed by 151 images corresponding to three prehispanic monoliths, searching to reconstruct in 3D each one of them. From those 151 images 91 were selected for the reconstruction process. According to this, 3 data subsets are conformed and distributed in this way: The first one has 26 images, the second one is integrated by 19 images and the third one contains 46 images, where each subset belongs to one monolith.

III. Algorithm

A. Image Segmentation

Even though what could be designated as the principal phase of the 3D models generation process from an image set lies in the work done inside the SfM pipeline, to this phase underlies the one that results determinant in the visual quality of the generated model, as will be exposed next up in the results analysis section. It is indispensable the definition of regions of interest for the dense point cloud generation. In the present approach this was made through image segmentation techniques which even though make this process easier and faster, they are not fully automatic. It was considered relevant to do not perform this phase in an automatic way seeking to give the user, the possibility of freely choose which part specifically wishes to reconstruct. Based on this, the employed technique was the interactive graphcuts [4] which consist in the pipeline presented in Fig. 2.

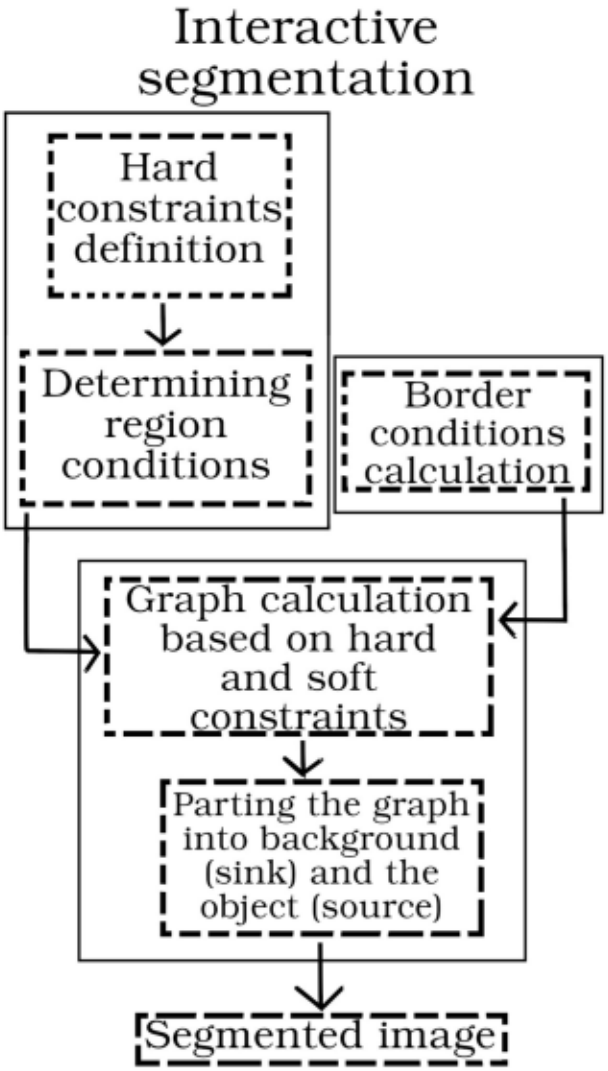


Fig.2: Interactive image segmentation workflow.

This technique consists in the definition of a series of hard constraints on the user side and a set of soft constraints which are inherent to each image. Starting from these constraints a graph is generated and it has as nodes each one of the image pixels, and a pair more of nodes referring to the interest object and the background of the image the user seeks to segment. Each node is interconnected weightily, in such way that when the cut through the graph is done, the result is the segmented image according to the probability of each pixel to belong to the object or to the background.

The graph mentioned above consists of two principal nodes, the first of those nodes corresponds to the interest objects in the image (object) and the second one to those pixels the user wishes to take away from the region of interest (background). Initially, all the pixels are connected both to the object node (usually named as source) and the background node (usually named as sink). Additionally the n-neighborhood of each pixel is considered to establish the existence of edges inside the image and consolidate the soft constraints that will be used in the graph cut phase. These borders can be found through gradient and Laplacian techniques [4].

Based on above, it is known that each pixel of the image has a probability (or a weight) in its connections to sink and source nodes, at the same time holds relations with its n-neighborhood; And finally with this information a cut is done through the connections of every pixel, assigning it to the object of interest or the discarded region.

B. Structure from motion pipeline

This consists of a series of computational algorithms and processes for the point clouds generation from the overlapping of a set of images captured from multiple points of view of one same scene. The Structure from Motion technique works initially in the same way than traditional photogrammetry techniques, although it differs in not needing to know previously the 3D key locations to do the reconstruction [5].

The whole Structure from Motion pipeline, is described in Fig. 3 and the first procedure which is done consists in finding the key points of the image set. A key point is a pixel or group of pixels that are easily recognizable in a pair of images of the same scene [6]. To find this set of key points in each image, the Scale-invariant feature transform [7] algorithms are used.

Basically, SIFT applies in first place a Gaussian difference to identify a set of sections in the image that are determined subtracting to the original image, a Gaussian blurred version of itself. Once these regions are defined, it proceeds to search features (points) of the image, these points are those which remain unalterable to local transformations applied to itself (such as rotations, changes in the illumination conditions and certain kinds of scale changes). The above procedure allows to find a set of key points where the descriptors are calculated for the comparison of features between multiple images. Indeed, it is to search for common features between each pair of images having a key point, so in an iterative way the correspondences with the other images are found and in this way the sparse point cloud will be built.

It must be highlighted that the amount of descriptors will be related to the illumination, the surface texture and the spatial resolution of the picture (The closer the camera is posed to the feature of interest, the bigger spatial resolution is obtained, and consequently a higher and accurately resolution in the sparse point cloud). Once the descriptors has been assigned, it proceeds to build the sparse point cloud, and for this are employed algorithms like RANSAC [8] and the nearest neighbor approximation [9]. For it at least two key points (descriptors) and three images must be compared; This means to search similarities between descriptors so if there are nonstatic elements in the scene (such as people or animals), these will be automatically removed from the point cloud.

The procedure to project in a 3D space the key points implies a re-projection error and this consists in the euclidean distance between the projected point and the calculated one. This error is defined by the expression:

$$\sum (P, X) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left\| u_{ij} - P_i X_j \right\|^2$$

Where X_j is the projected 3D point of the correspondence j, while u_{ij} is the correspondence between the points and P_j is the projection matrix of the camera for the image i. This projection matrix depends on the intrinsic parameters (Focal length on the x axis, focal length on the y axis, principal point and optical center) of the camera employed in the data capture phase. The process to reduce the re-projection error is known as bundle adjustment.

The Clustering View for Multi-view Stereo (CMVS) algorithms are employed and their function is to take as input the resulting data from the bundle adjustment and distribute this entry in many data arrays with such a size that allows to optimize the administration of this data [10]. To proceed with the 3d reconstruction process of each one of those arrays, Patch-based Multi-view Stereo algorithms are used, and it is through this algorithm that a dense point cloud is generated and will eventually be used in the generation of a polygonal mesh.

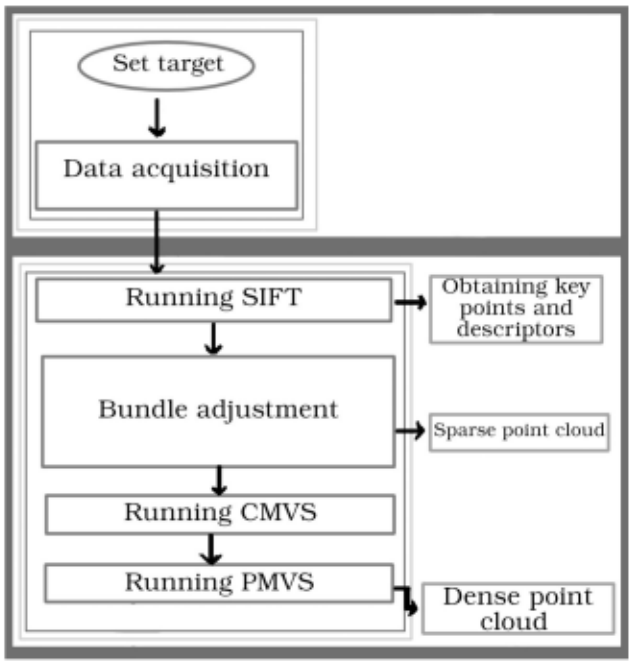


Fig. 3: Structure from Motion algorithm for the dense point cloud generation

Following the development and improvement of this technique, multiple applications and software specialized in the production of dense point clouds have emerged. Actually this kind of procedure can also be done through free and open source software (FOSS) as Open SfM, which is an open source library to generate dense point clouds from pictures.

C. Triangular meshes generation

The OpenSfM pipeline generates a dense point cloud from overlapped images, even though to ease the observation of the reconstructed object and to filter noisy information present in the point cloud (Fig. 4), it is proposed a workflow which searches to generate a triangular mesh of the dense point cloud.

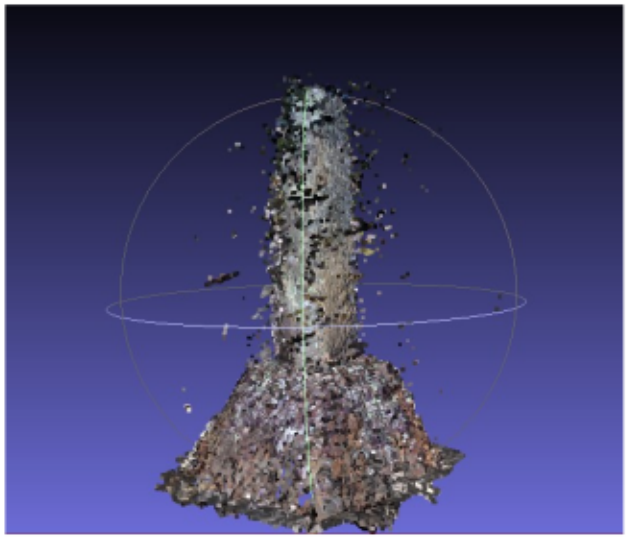


Fig. 4: The generated dense point cloud through the SfM algorithm presents noisy information and the goal is to remove this noise.

This workflow consists of making, in first place, a manual cleaning procedure upon the vertex set generated before. The purpose is to delete information (vertices) which are foreign to the surface of the reconstructed object. Once the manual cleaning is done, a sub-sampling procedure is executed on the point cloud, and this is done through the method presented in [11]. Finally it proceeds to generate a triangular mesh (Fig. 5) using the approach proposed by [12], for the archaeological reconstructions generation it is suggested to use a low or medium weighing, and increase (If there is still noise in the point cloud) the minimum samples value.

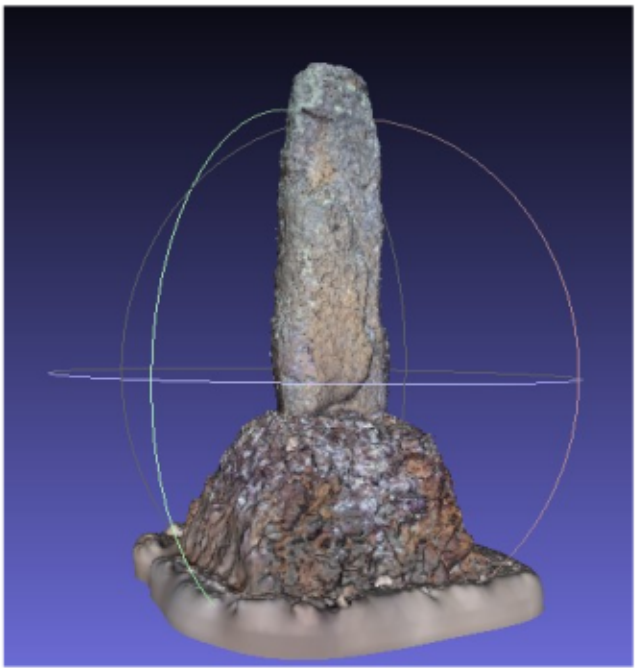


Fig. 5: Triangular mesh generated after the cleaning and sub-sampling procedures.

TABLE I. Amount of vertices in the point cloud and triangular mesh for the (a), (b) and (c) monoliths

	(a)	(b)	(c)
Vertices in point cloud without binary masks	1187874	645377	1646425
Vertices in point cloud with binary masks	284145	44040	243130
Vertices in triangular mesh without binary masks	190697	161157	170019
Vertices in triangular mesh with binary masks	265245	41982	173335

IV. Results

Six triangular meshes that correspond to the reconstruction of three stone monoliths belonging to the muisca culture, and located in the well of Hunzahúa in Tunja, Boyacá, Colombia were obtained. To reach these polygonal meshes, the procedure proposed in this article is: Binary masks generation, dense point cloud production and consolidation of the triangular mesh. Fig. 6 shows the three real objects (a), (b) and (c), their dense point clouds and the polygonal meshes generated using or not the produced binary masks.

The visual repercussions of using or not the binary masks for the reconstruction of outdoors was analyzed. Each studied object has a dense point cloud and a triangular mesh generated without employing binary masks, and a dense point cloud and a triangular mesh generated using binary masks to delimit regions of interest.

Table I shows the statistics of the faces and vertices amount for the (a), (b) and (c) objects, the first two rows correspond with the number of vertices produced along OpenSfM pipeline, so the third and fourth rows are the quantity of vertices after the sub-sampling procedure. The total of vertices removed through the process to reach a triangular mesh is counted. This value can be appreciated in table II. For the generation of triangular meshes the number of minimum samples was three, while the weight of the interpolation was zero.

The produced triangular meshes preserve details like reliefs and colors from the respective original object (see Fig. 6), nevertheless the results employing binary masks and do not using them differ both visually and computationally.

The generation and implementation of regions of interest for the dense point cloud production implies a reduction in the noise amount (foreign vertices) of the generated clouds as shows Table II, because the average estimated percent of noise has a value of 82.5675% for the reconstructions without binary masks implementation, while the reconstructions where the binary masks are implemented have an average of 12.7748% noisy information.

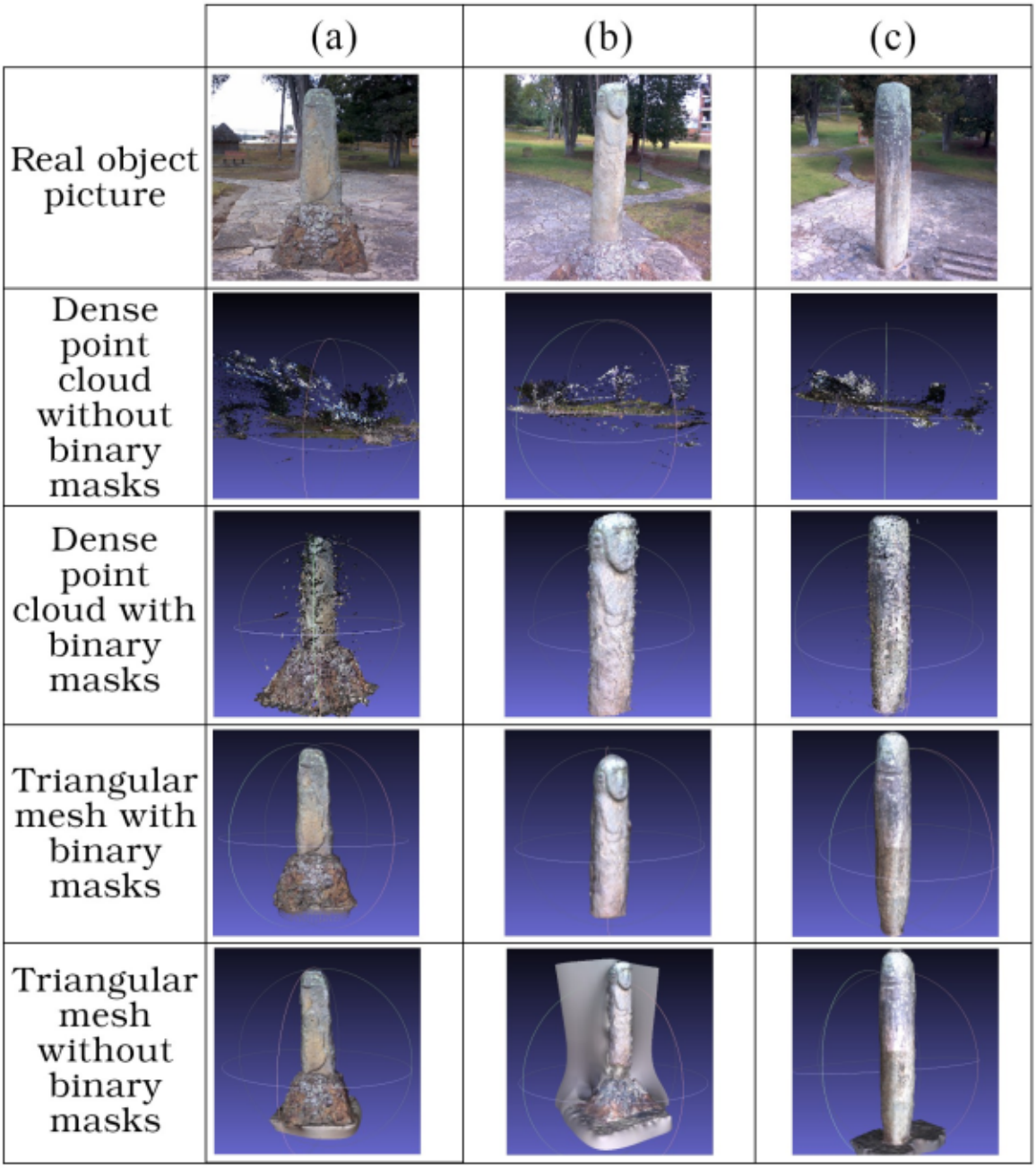


Fig. 6: Visual comparison of the generated dense point clouds and their respective 3D models, using and without using binary masks.

TABLE II. Percentage of removed vertices in each of the reconstructions, with and without employing binary masks.

	(a)	(b)	(c)	mean
Percentage of removed vertices without binary masks	83%	75.029%	89.6735%	82.5675%
Percentage of removed vertices with binary masks	6.6516%	4.6730%	27%	12.7748%

The 'mean' column represents the mean percentage of vertices removed in the cleaning and sub-sampling procedure.

V. Conclusions

A robust methodology for the production of triangular meshes from objects and archaeological contexts was proposed, this is developed with low cost devices and under the free and open source software framework. This methodology is sensitive to the interpretative component that integrates the archaeological discipline, and in that way this approach seeks to respect the autonomy of the user when the binary masks are generated and the first phase of the sub-sampling process is performed in a semiautomatic way.

Is problematic search to establish quantitative frameworks for the evaluation of the generated models because generally this archaeological objects neither have regular surfaces or are objects which have been recently discovered and this means that rarely have a reliable 3D reconstruction so the task of determining which generated model is the most reliable one, could be related directly to the person who makes the 3D reconstruction. This task consists into establish comparisons between the real object and the generated triangular meshes, searching to notice which model preserves better the details (reliefs, colors, protuberances and proportions) of the real object. A quantitative evaluation could lie in the establishment of many ground control points while the data acquisition phase, to take the triangular mesh to the same coordinate system than the real object and from this point, define if the reconstructed object has the same metrics than the real one.

It is necessary to build robust data sets (large amounts of images) when trying to increase the detail level of the point cloud, and, consequently of the triangular mesh. It is important also to build the data sets with both kinds of images, images containing the entire object and images in detail of the reliefs of the real object surface. That is why the virtual reconstruction of (b) is the most precarious one.

The noise is generated because of the presence of elements like trees and clouds, this implies a greater need of computational resources to show the reconstruction and of course an additional difficulty to visualize the reconstructed object.

ACKNOWLEDGMENT

This work was supported by the Vicerrectoria de Investigaciones Universidad Militar Nueva Granada through the PIC-ING-3531 project.

References

[1] F.Carvajal-Ramírez, A.D.Navarro-Ortega, F.Agüera-Vega, P. Martínez-Carricondo, and F. Mancini, "Virtual reconstruction of damaged archaeological sites based on unmanned aerial vehicle photogrammetry and 3d modelling. study case of a southeastern iberia production area in the bronze age," Measurement, vol. 136, pp.225–236, mar 2019.

[2] M. A. Quintana Rosales, "Registro de una secuencia temporal de nubes de puntos utilizando tecnología kinect para la reconstrucción tridimensional de material arqueológico," Pontificia Universidad Católica del Perú, Lima, Perú, tech report, Jun. 2015.

[3] N. Lercari, "3d visualization and reflexive archaeology: A virtual reconstruction of Çatalhöyük history houses," Digital Applications in Archaeology and Cultural Heritage, vol. 6, pp. 10–17, sep 2017.

[4] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in n-d images," in Proceedings of "Internation Conference on Computer Vision", Vancouver, Canada, July 2001 vol. I, p.105, 2001.

[5] M. Westoby, J. Brasington, N. Glasser, M. Hambrey, and J. Reynolds,"'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications," Geomorphology, vol. 179, pp. 300–314, dec2012.

[6] J. Meza, A. G. Marrugo, G. Ospina, M. Guerrero, and L. A.Romero, "A structure-from-motion pipeline for generating digital elevation models for surface-runoff analysis," Journal of Physics: Conference Series, vol. 1247, p. 012039, jun 2019. [Online]. Available:https://doi.org/10.1088/1742-6596/1247/1/012039

[7] D. Lowe, "Object recognition from local scale-invariant features," in Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157 vol.2.

[8] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol. 24, no. 6, p. 381–395, Jun. 1981. [Online]. Available: https://doi.org/10.1145/358669.358692

[9] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y.Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," J. ACM, vol. 45, no. 6, p. 891–923, Nov. 1998. [Online]. Available: https://doi.org/10.1145/293347.293348

[10] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis,"IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 32, no. 8, pp. 1362–1376, 2010.

[11] M. Corsini, P. Cignoni, and R. Scopigno, "Efficient and flexible sampling with blue noise properties of triangular meshes," IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 6, pp. 914–924, 2012.

[12] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," ACM Trans. Graph., vol. 32, no. 3, Jul. 2013. [Online]. Available:https://doi.org/10.1145/2487228.2487237