Processing 3D panoramic photos into point cloud modell

Regina Hanna BALATON
Institute of Geoinformatics
Óbuda University, Alba Regia Technical Faculty
Székesfehérvár, HUNGARY
reginahannabalaton@gmail.com

Gergely LÁSZLÓ
Institute of Geoinformatics
Óbuda University, Alba Regia Technical Faculty
Székesfehérvár, HUNGARY
laszlo.gergely@amk.uni-obuda

Zoltán TÓTH
Institute of Geoinformatics
Óbuda University, Alba Regia Technical Faculty
Székesfehérvár, HUNGARY
toth.zoltan@amk.uni-obuda

***Abstract* — 3D modeling based on traditional photographs is a fairly long-used and widespread method in photogrammetry. In this article, we present the process of model production based on 360° panoramic images, describing the problems, advantages, and disadvantages of the procedure. We use a Ricoh Theta Z1 camera for photos, and Agisoft Metashape software for processing.**

Keywords — photogrammetry, panoramic images, 3D modeling

I. INTRODUCTION

Agisoft Metashape is a stand-alone software product that performs photogrammetric processing of digital images and generates 3D spatial data to be used in GIS applications, cultural heritage documentation, and visual effects production as well as for indirect measurements of objects of various scales. [1]

1. Figure - Ricoh Theta Z1 camera

The RICOH THETA Z1 supports approximately 23-megapixels resolution (6720 x 3360 pixels) 360-degree still image photo shooting as a high-end model of the RICOH THETA camera series, which can shoot spherical images in a single shot, giving consumers the opportunity to adventure, discover and capture life in 360. [2]

A three-dimensional survey of building interiors may be necessary more and more often, as designers and operators of registers have realized on the one hand that in many situations it is not possible to represent objects in a complex way in 2D, and on the other hand, BIM (building information modeling), which is a procedure that includes the creation and management of digital spatial models of physical and functional properties of places.

To satisfy the spatial information needs of these systems, there are several survey methods, starting from traditional total station surveys, 3D laser scanning and terrestrial photogrammetry. In a different way than these, but in each case, we get a final result in which there are points with x, y, z coordinates and some quality information. There are several differences between the procedures given in the above example, one of which is the absolutely not negligible asset investment aspect. Of the three, the tool for the photogrammetry process examined in the article is the cheapest, so this was one of the aspects of our choice.

II. TERRESTRIAL PHOTOGRAMMETRY

The branch of science dealing with the preparation and processing of measurement images and photographs designed for measurement and evaluation purposes from the ground position is called terrestrial photogrammetry. It can be used where an unobstructed view of the area, object, or surface to be measured is ensured. Nowadays, its application in cartography is beginning to come to the fore again with the spread of 3D databases, previously it was only used in those specialist areas where not only the processing of metric data is important, but also the additional information provided by the photograph.

III. CHARACTERISTICS OF TERRESTRIAL PHOTOGRAMMETRY

- it is not necessary to approach the examined object or facility during the measurement (this is why terrestrial photogrammetry is also a type of remote sensing),
- the time required for the field measurement is short,
- due to the short exposure (exposure) time, it is also suitable for measuring fast processes,

- the measurement images contain a lot of information (quantitative and qualitative),
- the evaluation is fast and can be easily automated.

IV. RECORDING ARRANGEMENTS OF TERRESTRIAL PHOTOGRAMMETRY

In terrestrial photogrammetry, we also speak of single-image or planar photogrammetry and two-image, or spatial photogrammetry. Recording layouts can also be grouped accordingly:

- one-shot procedure,
- spatial photogrammetry procedure.

From the point of view of this article, the spatial solution is important.

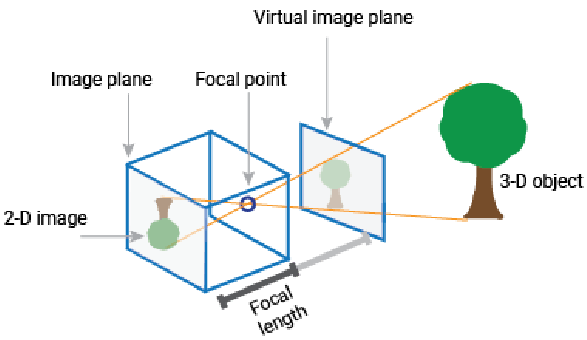
In spatial photogrammetry, we aim for the spatial (three-dimensional) determination of the points to be measured, for which we need to take at least two pictures of the object to be measured. In both pictures, the object to be measured must be on it. In previous, traditional procedures, the photographs taken for measurement purposes were taken from positions with known horizontal and height data, where the direction of the camera axes was also known.

Nowadays, the evaluation and model creation is done with a multi-image 3D reconstruction process. The essence of photographing is to depict spatial objects in a plane, this means that depth is lost, the process we use is the opposite of this.

In doing so, we need several images, since we cannot determine from a single image where a spatial point is located on the projection line, and which point on this line actually corresponds to the pixel. If we already have two images that depict the point in question from different locations, the position of a 3D point can already be determined as the intersection of the two projection rays. This process is known in the literature as triangulation. The key to the process is knowing the relationship between the recording locations, on the basis of which the appropriate transformation equations can be written.

Reconstruction is a multi-step process.

In the first step, we need to calibrate the camera, during which the internal and external orientation parameters are determined. Photogrammetric processing cannot be performed without this step.



2. Figure - Imaging Process

By internal parameters we mean the location of the optical center on the image plane, the focal length and the optical distortion parameters of the lens. Described in matrix form, they are:

$$K = \begin{bmatrix} a_x & \gamma & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The K contains 5 intrinsic parameters of the specific camera model. These parameters encompass focal length, image sensor format, and camera principal point. The parameters $a_x = f * m_x$ and $a_y = f * m_y$ represent focal length in terms of pixels, where m_x and m_y are the inverses of the width and height of a pixel on the projection plane and f is the focal length in terms of distance. γ represents the skew coefficient between the x and the y axis, and is often 0. u_0 and v_0 represent the principal point, which would be ideally in the center of the image.

Nonlinear intrinsic parameters such as lens distortion are also important although they cannot be included in the linear camera model described by the intrinsic parameter matrix. Many modern camera calibration algorithms estimate these intrinsic parameters as well in the form of non-linear optimization techniques. This is done in the form of optimizing the camera and distortion parameters in the form of what is generally known as bundle adjustment.

External orientation parameters

$$\begin{bmatrix} R_{3x3} & T_{3x1} \\ 0_{1x3} & 1 \end{bmatrix}_{4x4}$$

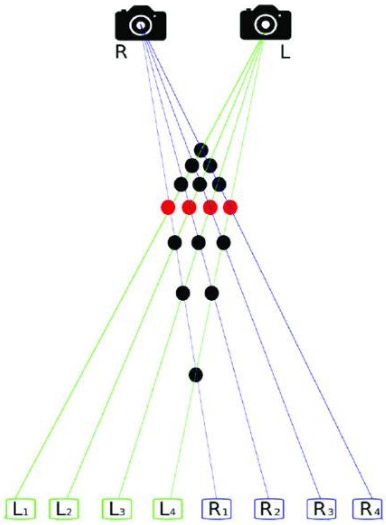
R and T are the external orientation parameters that denote the transformations of the coordinate system from the coordinates of the 3D world to the coordinates of the 3D camera. Accordingly, the external parameters determine the position of the center of the camera and its orientation in world coordinates.

T is the position of the origin of the world coordinate system given in the camera-centered coordinate system. The position of the camera C can be expressed in world coordinates based on the following formula: [3]

$$C = -R^{-1}T = -R^{T}T \ (since\ R\ is\ a\ rotation\ matrix)$$

Determining the depth of field

Determining depth of field is the most challenging part of the process, as this is where the important spatial information missing from images is determined - depth. The correspondence problem, finding the match between the two examined images, thus the position of the matched elements in 3D space, is the question to be answered.



3. Figure – Restore depth of field

Mesh creation

Once we have the multiple depth maps, we need to combine them to create the final mesh by calculating the depth and projecting it from the camera. Camera calibration will be used to identify where the many meshes created by depth maps can be combined to create a larger mesh that already provides viewing from multiple viewpoints.

Coloring the mesh model

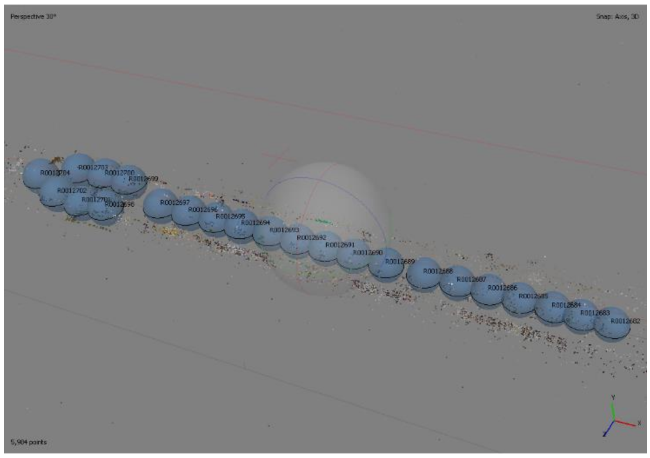
In this step we have a full 3D mesh, which may be the ultimate goal, but usually we want to apply the color of the original photos to the mesh. This can range from randomly projecting images onto the mesh for super-resolution, to combining textures, to segmenting the mesh by material, including things like reflections and diffuse properties.

V. Presentation of the Survey

The survey was carried out with a Ricoh Theta Z1 360° camera, which we borrowed from the Slovak University of Technology in Bratislava. The surveyed area was the ground-floor corridor of the Pirosalma street building of the Alba Regia Technical Faculty of Óbuda University.

4. Figure - The test area

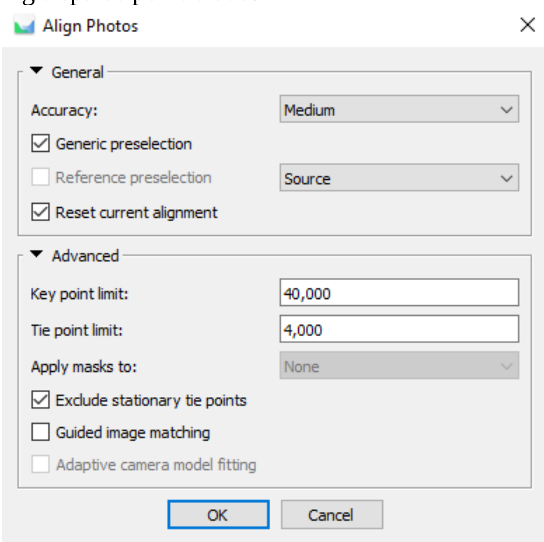
On this occasion, we did not use the point signals provided by Matterport for the survey, as we did not want to do the processing with its own program, so that we could export a point cloud. A total of 23 images were taken with appropriate overlap in the arrangement shown in the image below:



5. Figure – The places of the pictures

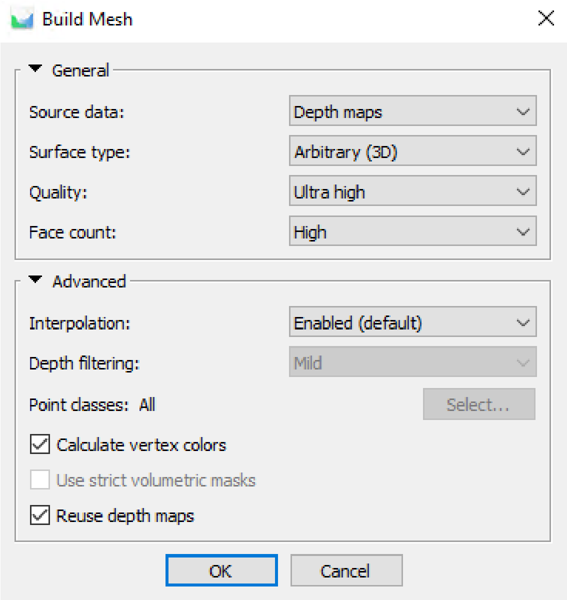
Processing

The processing was done with the Agisoft program. First, we imported the saved JPEG format images into a new project, then set the camera properties to use a spherical camera model for processing. The first step of the processing was the relative alignment of the photographs performed at the same time as the calibration. This is done in such a way that the program searches for identical points on each pair of images, thereby creating a sparse point cloud.



6. Figure - Align Photos properties

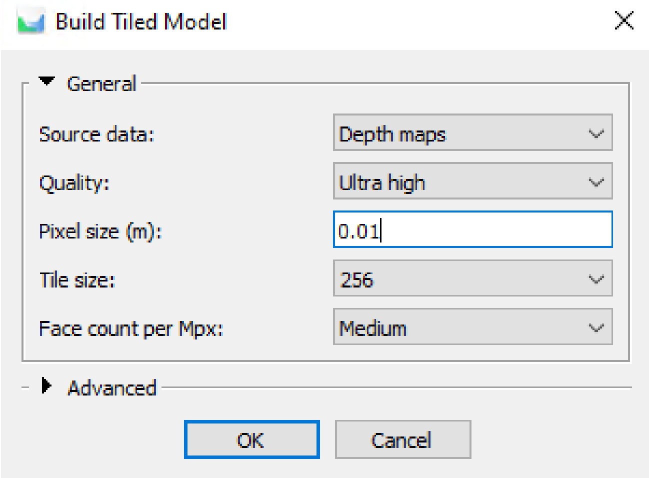
The second step, as previously described, was to define the depth-of-field map, thereby restoring the 3D nature of the surveyed data. With this, the creation of the mesh model is also done in one function, as the first 3D visualized final product.



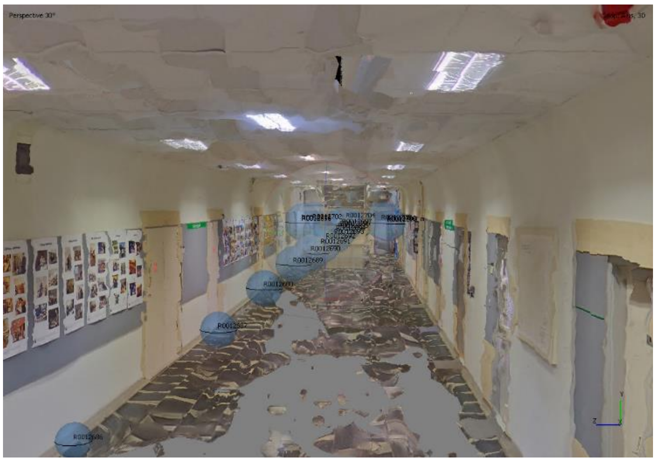
7. Figure - Mesh properties

8. Figure - Mesh Model

During the third, optional step, a tiled model can be made from the mesh model, which can be exported to formats supporting 3D models for further processing with external software.

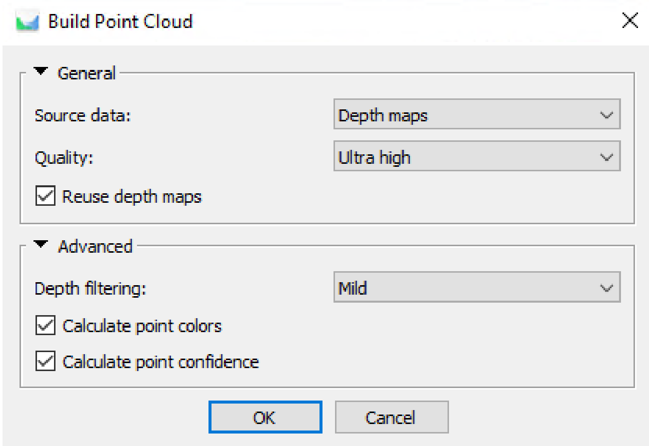


9. Figure - Tiled Model properties

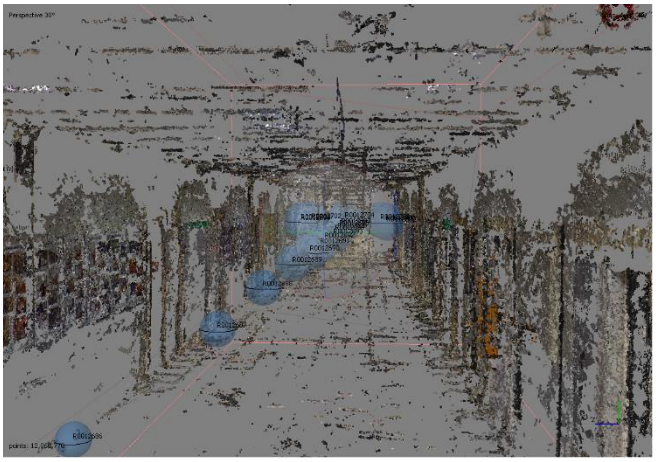


10. Figure - Tiled Model

In the last step, the 3D point cloud is created, using the results created in the first two phases. The pictures below show the settings and the final product:



11. Figure - Point Cloud properties



12. Figure - Point Cloud

VI. DISCUSSION

The images of the results show that the method has weaknesses. First of all, there will be no spatial information about visually homogeneous surfaces, such as the floor, wall and ceiling, because the software and the procedure it uses cannot calculate spatial intersection points, so there could only be a point formed on those parts if we interpolated a plane to those places, which would negatively affect accuracy. Another obvious problem is that dots are formed in places where they obviously shouldn't be. These can be caused by reflections or camera distortion. Their repair could be solved by manual cleaning or calibration.

13. Figure - Generated noise below floor level

VII. Conclusion

In the case of images taken on inhomogeneous surfaces, a point cloud of adequate density can be created for the entire area, but under average conditions, you have to account for parts with missing data. By solving the problem of camera calibration, the points formed on the edges of the camera would improve in terms of accuracy. In any case, geometrical validation would be necessary in comparison with other measurement methods.

Acknowledgment

The research was supported by the New National Excellence Program 2023 (ÚNKP 2023). In addition, we are grateful to the Slovak University of Technology in Bratislava for the loan of the RICOH THETA Z1 360° camera, which is essential for the realization, without which the survey of the research could not have been realized.

We would like to express our special thanks to Ján Erdélyi and Richard Honti, who gave us a quick and professional guide regarding the use of the camera and the processing.

Új Nemzeti Kiválóság Program

National Research, Development and Innovation Office Hungary

Kulturális és Innovációs Minisztérium

References

[1] https://www.agisoft.com/

[2] https://ricohtheta.eu/products/ricoh-theta-z1

[3] B. Triggs; P. McLauchlan; R. Hartley; A. Fitzgibbon (1999). "Bundle Adjustment — A Modern Synthesis" (PDF). ICCV '99: Proceedings of the International Workshop on Vision Algorithms. Springer-Verlag. pp. 298–372. doi:10.1007/3-540-44480-7_21. ISBN 3-540-67973-1.

[4] Image and depth from a conventional camera with a coded aperture Anat Levin, Rob Fergus, Fredo Durand, William T. Freeman, MIT

[5] Martinello, Manuel; Favaro, Paolo (2012). "Depth estimation from a video sequence with moving and deformable objects". IET Conference on Image Processing (IPR 2012) (PDF). p. 131. doi:10.1049/cp.2012.0425. ISBN 978-1-84919-632-1.

[6] https://www.nive.hu/Downloads/Szakkepzesi_dokumentumok/Bemeneti_kompetenciak_meresi_ertekelesi_eszkozrendszerenek_kialakitasa/20_2241_004_100915.pdf