IMAGE RETRIEVAL FOR 3D MODELLING OF ARCHITECTURE USING AI AND PHOTOGRAMMETRY

F. Condorelli *

Free University of Bozen, Faculty of Education, Brixen, Italy – francesca.condorelli@unibz.it

KEY WORDS: Image generation, Photogrammetry, Artificial Intelligence, Cultural Heritage, 3D Modelling

ABSTRACT:

This research is intended to provide an initial solution to the problem of finding images for processing by photogrammetry in special cases where these do not exist. An overview of existing artificial intelligence-based algorithms that enable the extension of source image dataset is reported. In particular, this research focused on the use of prompt-to-image systems for obtaining images to be used in reconstruction and then in the next step of 3D modelling. Thus, the combined use of these three techniques, AI, photogrammetry, and modelling allowed the creation of a model of a building that never existed except in the collective imagination, which is the tower of Babel. In particular, the case study chosen is the illustration in Kircher book present in the library of the Brixen seminary that is closed to the public and for which it was necessary to create a tool to enhance the value and knowledge of this heritage for external users. Therefore, the creation of an augmented reality app enabled the visualization of the model created by offering possibilities for immersive experiences and dissemination of the research to a wide audience.

1. INTRODUCTION

Recent research into the use of Artificial Intelligence (AI) applied to 3D modelling has focused on the creation of surfaces and meshes from a limited amount of data. This is because it is very often the case that a very limited dataset is available for various reasons. Especially when 3D reconstruction is to be obtained using the technique of photogrammetry, in some cases it is difficult to meet the minimum requirements for processing.
This study is therefore part of this line of research, dealing with the problem of how to extend the image dataset in order to improve 3D modelling. In particular, the experiments focused on the study of illustrated architectures in order to assess how useful the technique used can be extended to images of real cultural heritage.
In order to achieve this objective, a combination of two techniques, photogrammetry and AI, are used for image retrieval with the aim of improve the 3D modelling of architecture. Photogrammetry involves using multiple photographs of a building or structure to create a 3D model. AI algorithms can be utilized to assist with image processing and help automate the photogrammetry process.
Expanding an image dataset for photogrammetric reconstruction involves using AI algorithms to generate additional images or data to complement existing image data. The goal is to increase the quantity and quality of the image data available for photogrammetric reconstruction, resulting in more accurate and detailed 3D models.
This paper discusses the creation of a 3D model of an illustrated architecture that therefore never existed. Therefore, photogrammetry is applied as a technique that can help the standard modelling process. Metric analysis is therefore omitted in this first phase of the research. The innovation then is to use artificial intelligence-based techniques for dataset expansion useful for both manual modelling and with photogrammetry. The resulting 3D model is used at this stage mainly for visualization and scientific dissemination purposes of the research results.

2. AI BASED IMAGE GENERATION ALGORITHMS: A STATE OF THE ART

Some ways AI can be used to expand the image dataset include for example Image Synthesis in which algorithms can be used to generate synthetic images that complement existing real images. These synthetic images can be used to fill in gaps or provide additional viewpoints that would otherwise be difficult or impossible to obtain. Another technique is Image Enhancement in which algorithms can be used to enhance the quality of existing images, such as by reducing noise or improving the resolution, which can lead to better results in photogrammetric reconstruction. Also, Image Augmentation can be used in which algorithms allow the augmentation of existing images, such as by adding variation in viewpoint, illumination, or other parameters, to increase the diversity of the image data available for photogrammetric reconstruction.
By expanding the image dataset using AI, photogrammetric reconstruction can be improved, resulting in more accurate and detailed 3D models of buildings and structures.
Previous studies explored different kind of AI based algorithms for the retrieval of images. For example, Visual search engine algorithms are computer systems that use AI techniques to search for and retrieve images based on visual content, rather than text-based keywords. These algorithms use computer vision techniques to analyze the visual content of an image and compare it to other images in a database, to determine which images are most similar.
Some common AI algorithms used in visual search engine algorithms include for example Convolutional Neural Networks (CNNs) that are deep learning algorithms trained on large datasets of images to learn the visual patterns and features of images. They can be used to perform image classification and similarity comparison. Another type of algorithms are Image Embeddings, mathematical representations of images that capture their visual content. Image embeddings can be generated using deep learning algorithms, and they can be compared to other image embeddings to determine the

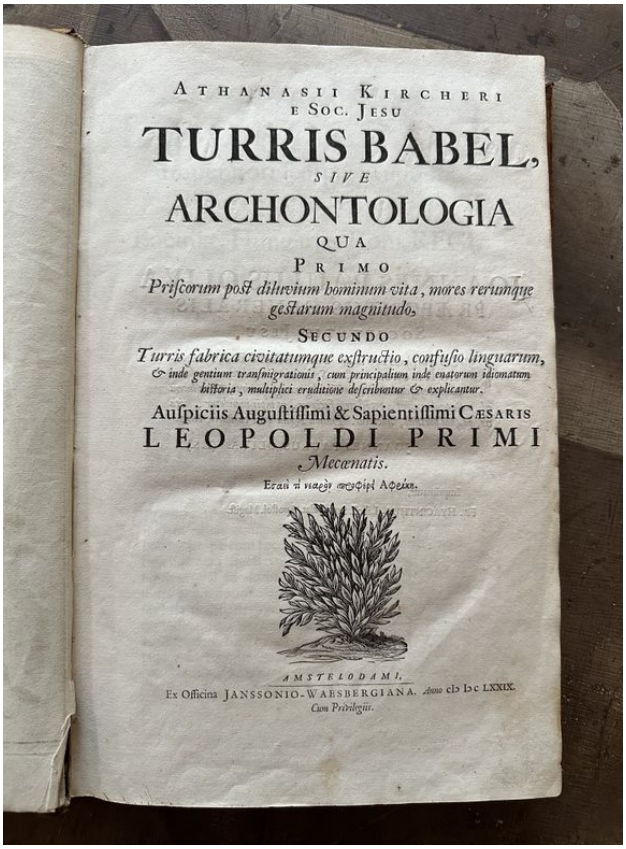
* Corresponding author

similarity between images. Finally, Local Feature Descriptors are algorithms that identify and describe the unique visual features of an image, such as corners, edges, or textures. They can be used to compare images and determine their similarity.
These AI algorithms can be combined and used in different ways to build effective visual search engine algorithms. By using AI to analyse the visual content of images, visual search engines can provide more accurate and relevant results compared to traditional text-based search engines.
Diffusion-based methods for image generation have recently been very successful. They involve using mathematical diffusion processes to generate new images from existing ones. These methods are based on the idea of spreading information or features from one image to another in a smooth and continuous manner. One popular diffusion-based method for image generation is called "image analogies." This method uses a source image and a target image, and maps features or information from the source image to the target image. The result is a new image that is similar to both the source and target images. Another diffusion-based method for image generation is called "image inpainting." This method involves filling in missing or corrupted parts of an image using information from surrounding areas. The diffusion process spreads information from the surrounding areas to the missing parts, effectively generating a new and complete image. Diffusion-based methods for image generation are useful for a variety of tasks, including image completion, super-resolution, and style transfer. By leveraging the mathematical diffusion process, these methods can generate new images that are visually consistent and preserve important features or information from the original images.
Finally other types of algorithms consist in AI based prompt-to-image systems. The emergence of AI solutions that can generate images from text queries cannot be attributed to a single strand of research but is rather the result of the convergence of different research efforts that have produced applications such as Midjourney (Muppalla et al. 2022).
In this research, these were chosen to be used for the automatic generation of images similar to the illustration under consideration to aid the photogrammetric and modelling process.

3. CASE STUDY AND METHODOLOGY

The case study chosen for this research is the illustration of Babel tower contained in the Jesuit Athanasius Kircher book (Figure 1 and Figure 2). An ancient copy of which is in the library of the Brixen seminary.

Figure 1. The Jesuit Athanasius Kircher book.



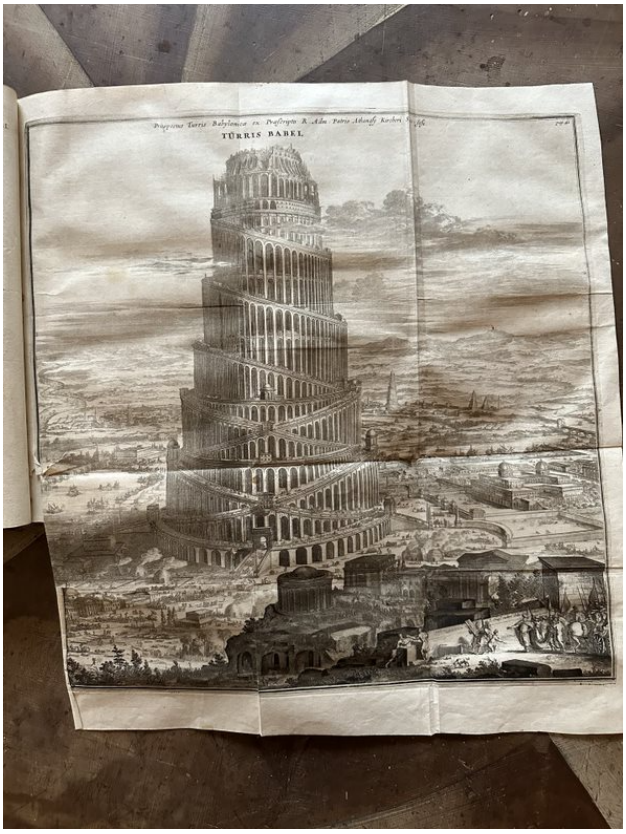


Figure 2. Illustration of the tower of Babel in the Jesuit Athanasius Kircher book.

In 1679, the Jesuit Athanasius Kircher published a volume in which he offered reflections linking the Tower of Babel to the

emergence of modern languages, beginning with the end of the universal Flood. The text, one of the Jesuit's many eclectic works, deals extensively with purely architectural issues, both through careful descriptions of ancient buildings and considerations of construction techniques, to demonstrate the possible untenability of the idea of building the tower. The presentation of Kircher's reasoning and descriptions is accompanied by a series of illustrations commissioned from Coenraet Decker and Gérard de Lairesse.

This case study is of particular interest for the following reasons: (i) it is an architecture, even if only imagined; (ii) it is a very famous symbol in different cultures and therefore easily known and replicable by AI; (iii) being in a book that is not easily accessible, as it is preserved in controlled conditions, it is important to find a way to make it known and show it to tourists and the scientific community, as suggested by UNESCO for the documentation and valorisation of tangible cultural heritage.

In order to show scholars and tourists what they cannot see because it is enclosed within the pages of the Athanasius Kircher's Archontologia, a workflow is proposed for the creation of the 3D model necessary for the visualization purpose. In fact, the methodology used was developed to solve the specific problem of the three-dimensional representation of a known architectural work with the help of some graphical representations made on the basis of the descriptions in the biblical writings.

This was made possible by implementing a three-part workflow, as shown in Figure 2.

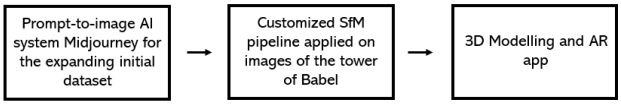


Figure 2. Proposed workflow.

The first part is the retrieval of images similar to the tower for expanding the initial database of images for photogrammetry using Midjouney. The second part concerns the reconstruction of the tower model. By combining the use of AI for image search and advanced photogrammetry, 3D modeling is improved. This is the most difficult part of the entire process, as it involves creating a model of an imaginary architecture that never existed. The third part of the process is the development of an augmented reality application to make the model of the tower navigable in 3D and accessible to the public.

Both the reconstruction of the tower and the implementation of the app were carried out with low-cost tools and using open source software.

3.1 AI based prompt-to-image system for the retrivial of images with Midjouney

AI based prompt-to-image systems such as Midjouney are web-based platforms where the user is asked to enter text and select from different images proposed by the system to refine the final result. The process consists of three steps. First, a text command is sent to a text encoder that has been trained to decode the text and assign numerical values to it. In the second step, a model called "prior" is used to associate the text encoding with the encoding of the corresponding image, which is assumed to be a visual manifestation of the same semantic information.

In terms of how images take shape, this process step is technically linked to the ability to deconstruct an organized set of specific data and then apply a reverse process that is able to reconfigure the data in a composition that is as close as possible to the original (Sohl-Dickstein et al. 2015). Thus, for digital images, the process involves altering and modifying the constituent pixels, whose precise position in digital space results in what we see, by introducing a known amount of noise (i.e., a set of randomly positioned pixels) that alters the image itself. The next step in this process is then to act in the opposite direction to reconstruct the original position of the pixels to obtain a new image that is similar to the original image (Rombach et al. 2022, Ho et al. 2020).

By repeating this process over large datasets, we can generate a predictive model that can compose images independently by rearranging a set of unstructured pixels.

What has been described so far allows us to generate images randomly: the next step, therefore, is to be able to incorporate into this chain a set of instructions that can guide the generation process in a predetermined, non-random direction.

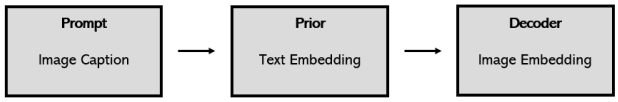


Figure 3. Workflow of AI based prompt-to-image system for the retrivial of images.

3.2 Creation of 3D model

Commercial software cannot process images of illustrations and of low numbers. This is because it does not allow customization of the algorithms underlying the processing, generating problems that lead to image alignment failure. For this reason, open source data processing software was used in this research, specifically COLMAP, an open source implementation of the Structure-from-Motion and Multi-View Stereo (MVS) algorithm developed by ETH Zurich (https://github.com/colmap/colmap, 2023).

The proposed workflow aims to extract the coordinates of specific measurement points in the retrieved images in order to use them as the basis for the 3D modelling phase.

The proposed workflow is shown in Figure 4.

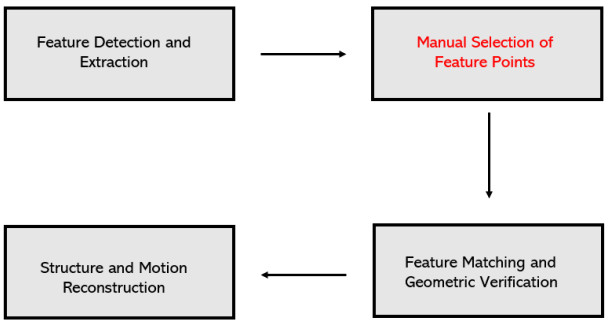


Figure 4. SfM workflow with the addition of 'Manual Selection of Feature Points' step.

The standard sequential SfM processing for iterative reconstruction in COLMAP is as follows: i) feature detection and extraction to find the feature points in the image; ii) Feature matching and geometric verification to find matches between feature points in different images; iii) Structure and motion reconstruction to obtain the final point cloud.

This standard pipeline was adapted [16] to extract the coordinates of specific points from the image dataset. The presence of some known specific points, selected by the human operator, allows the correct modelling of the object of study. After the feature detection and extraction phase, an additional

step of manual point selection (highlighted in red in Figure 4) was added to manually select feature points to be used in the subsequent feature matching and geometric verification phase. The algorithm for detecting and extracting new features from images in COLMAP is explained in detail in Condorelli et al. 2019. A summary is given here: with the standard feature detection and extraction step, COLMAP automatically detects the key points in the images, but it may miss an important radiometric angle in the image that also appears in other images. By introducing this step, it is possible to manually locate the characteristic point of interest and extract its 3D coordinates. The coordinates of the point of interest were measured using the WebPlotDigitizer tool (https://automeris.io/WebPlotDigitizer, 2023) and entered into the software. The data processed in COLMAP are stored in a user-defined database that can be easily managed. It is possible to enter new coordinates for characteristic points not already recognized by the automatic algorithm and therefore not present in the database. By creating a text file containing the image coordinates (x, y) expressed in pixels, as well as the scale and orientation information, and writing a line for each characteristic, it is possible to import known characteristics (e.g. individual points) into the database and use them in the matching phase. In this assisted processing, the most salient points such as corners and contours are selected and filtered out. The matching algorithm looks for the selected characteristic point in each image to estimate the equipotential line in the other images.

3.3 Visualization of the model with AR application

As for the third phase of the proposed workflow, the development of augmented reality is undoubtedly an excellent tool not only for visual communication as a knowledge tool, but also for effective interpretation of material cultural heritage. Markerless technology (Hammady et al., 2018) has been used in the development of the application to make the experience more immersive and user-friendly. Recent studies (Palma et al., 2018) have shown that this technology achieves the optimal levels of interaction with the user required for this type of experience (Shin et al., 2019).

4. RESULTS AND DISCUSSIONS

The implementation of Midjouney was useful to create new images of the tower. Using different prompts settings it was possible to recreate images similar to the structure of the tower in the illustration. In particular it was good to have a basis for the lateral view of the architecture suitable for the photogrammetric processing (Figure 5).
By detecting specific points in the photometric phase, 3D models of the tower were created based on the book illustration (Figure 6 and 7).

Figure 5. Examples of images of the tower create with Midjourney.

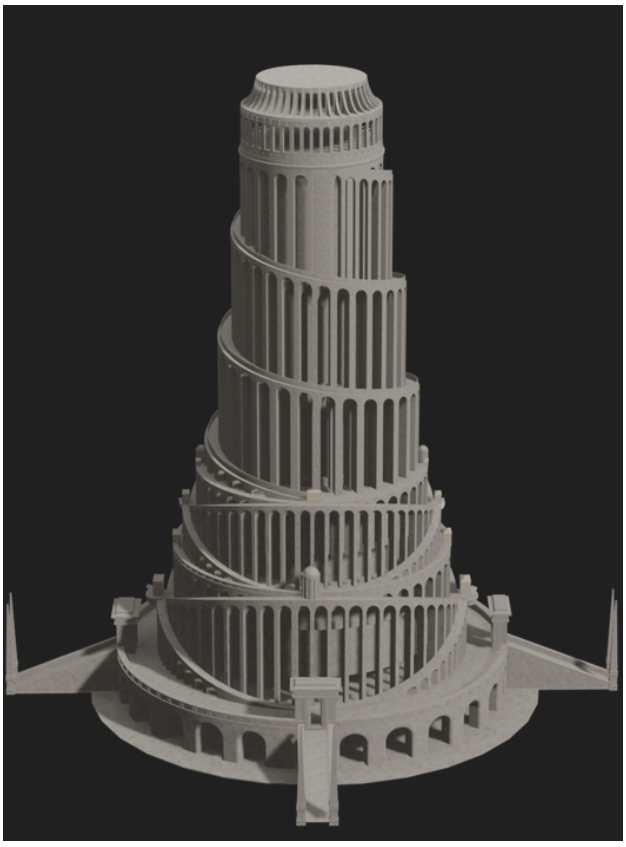


Figure 6. 3D model of the tower.

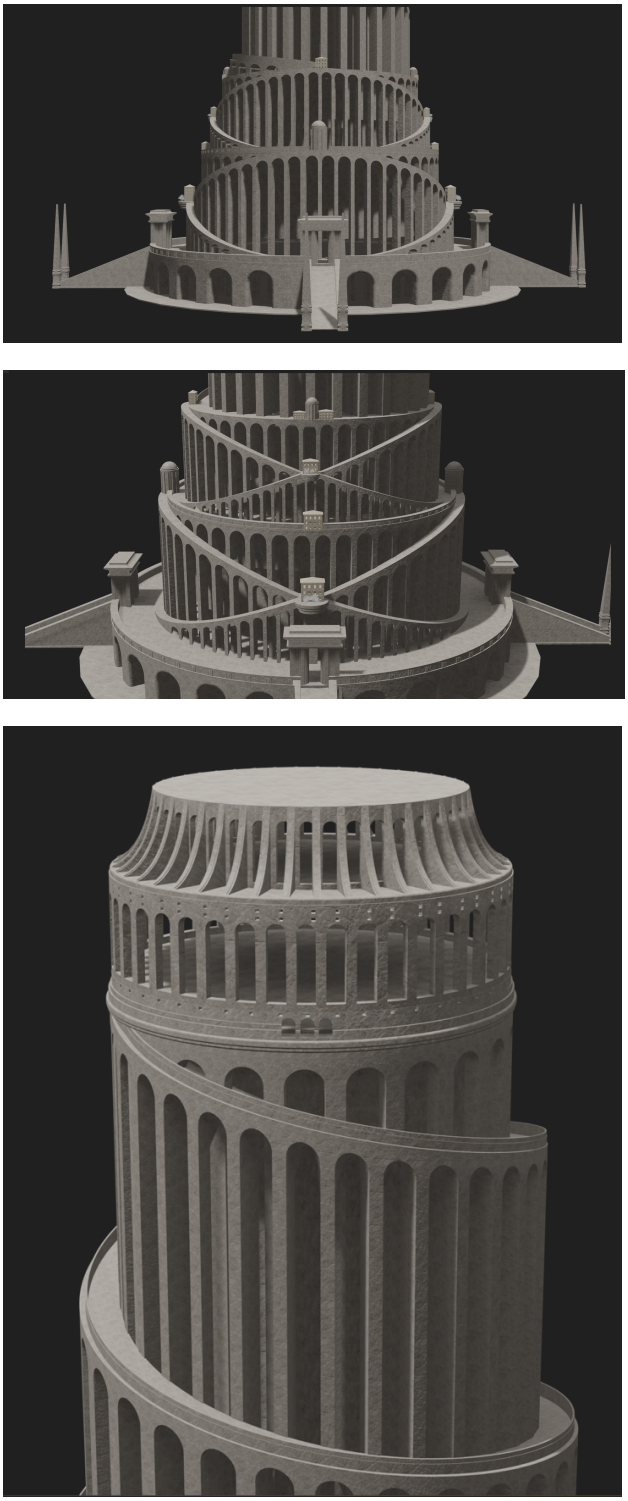


Figure 7. 3D model of the tower.

The results of the AR applications developed for the Athanasius Kircher Archontology are presented below. Thanks to the implementation of the Vuforia tool in the Unity environment, the application displays the navigable model of the tower in real time by simply pointing the device at the corresponding illustration in the book. The simplest and most commonly used type of marker in game applications is the image target, as shown in the Figure 8.

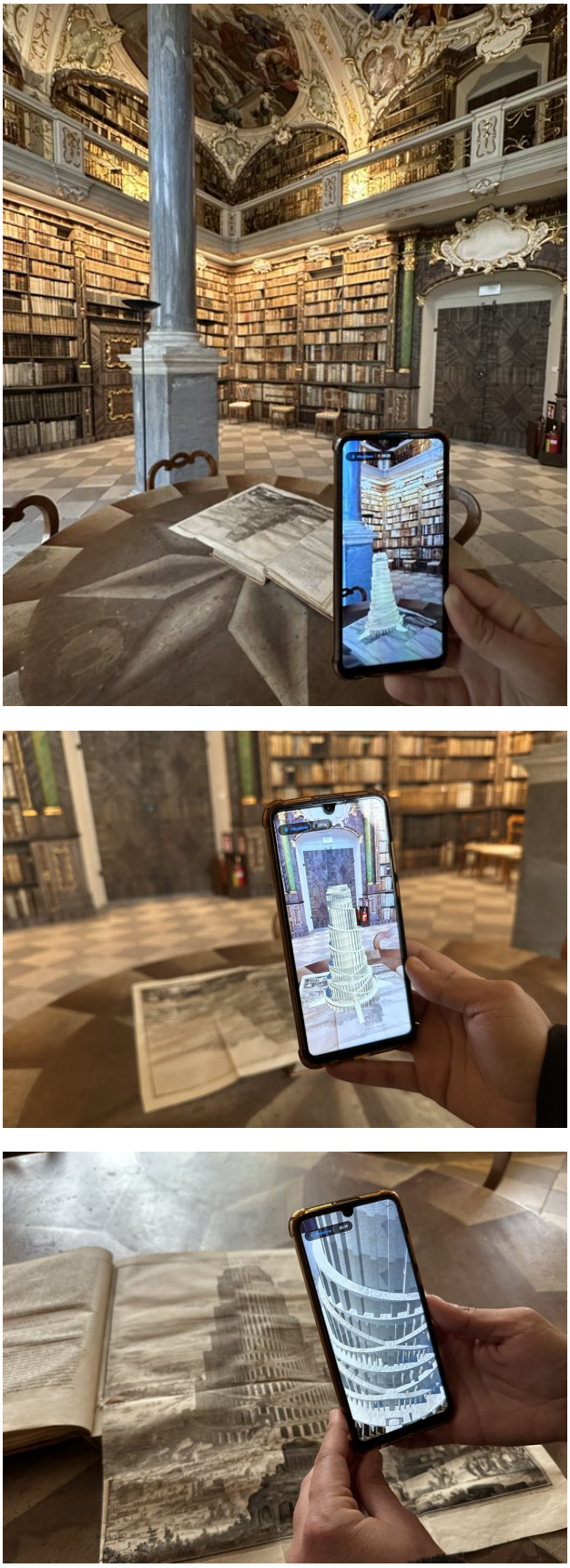


Figure 8. AR application for the visualization of the model of the tower.

5. CONCLUSIONS AND FUTURE WORKS

This paper presented a workflow that combines AI and photogrammetry to support 3D modeling of architecture in cases of particular difficulty in finding the initial image dataset. For this reason, it was chosen to reconstruct in 3D a building that is only imagined and not real, which is that of the tower of Babel featured in an illustration in the book by Kircher. AI and in particular the use of Midjourney was helpful in recreating images similar to the starting one, especially in creating the side views not present in the original drawing. Photogrammetry applied to this case study made it possible to support the standard modeling process by manually selecting specific points in the drawing that were useful in creating the model. The metric part related to photogrammetry was left out because there was insufficient data to obtain a metric analysis. Therefore, the model was used for visualization and research dissemination purposes to show an outside audience cultural heritage that is not accessible such as precisely the book under study. The advanced navigation device will be optimized, and the book content narration and illustration descriptions will be improved to ensure a smooth and attractive book experience. The proposed methodology may be repeated and tested on real architecture cases in such a way as to complete the workflow with metric analysis of the results. In this way, the resulting model can be used for other purposes and to take action on heritage.

REFERENCES

Borji, A., 2022: How good are deep models in understanding the generated images?, *Computer Vision and Pattern Recognition*, arXiv: 2208.10760

Borji, A., 2022: Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2, *Computer Vision and Pattern Recognition*, arXiv:2210.00586

Condorelli, F., Higuchi, R., Nasu, S., Rinaudo, F., and Sugawara, H.: Improving performance of feature extraction in SfM algorithms for 3D sparse point cloud., *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLII-2/W17, 101–106, https://doi.org/10.5194/isprs-archives-XLII-2-W17-101-2019, 2019.

Hammady, R., Ma, M., Powell, A., 2018: User Experience of Markerless Augmented Reality Applications in Cultural Heritage Museums: 'MuseumEye' as a Case Study. In: *De Paolis, L., Bourdot, P. (eds) Augmented Reality, Virtual Reality, and Computer Graphics*. AVR 2018. Lecture Notes in Computer Science(), vol 10851. Springer, Cham. https://doi.org/10.1007/978-3-319-95282-6_26

Ho, J., Jain, A., & Abbeel, P., 2020: Denoising diffusion probabilistic models. *In Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 6840-6851).

Palma, V., Spallone, R., Vitali, M., 2018: Digital Interactive Baroque Atria in Turin: A Project Aimed to Sharing and Enhancing Cultural Heritage. In*: Luigini, A. (eds) Proceedings of the 1st International and Interdisciplinary Conference on Digital Environments for Education, Arts and Heritage*. EARTH 2018.

Ploennigs, J., Berger, M., 2022: AI Art in Architecture, *Artificial Intelligence*, arXiv:2212.09399

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B., 2022: High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).

Shih, N.-J.; Diao, P.-H.; Chen, Y, 2019: ARTS, an AR Tourism System, for the Integration of 3D Scanning and Smartphone AR in *Cultural Heritage Tourism and Pedagogy*. Sensors 2019, 19, 3725. https://doi.org/10.3390/s19173725

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S., 2015: Deep Unsupervised Learning using Nonequi¬librium Thermody-namics. arXiv:1503.03585.