Multimodal Digitization Concepts for Acoustic Cultural Heritages

Using Photogrammetric Models for the Preservation and Auralization of Historical Acoustical Spaces

Dominik UKOLOV, Research Center for Digital Organology at the Museum for Musical Instruments of Leipzig University, Germany

Keywords: *photogrammetry — acoustics — sound — metaverse — auralization*

CHNT Reference: Ukolov, D. (2022). 'Multimodal Digitization Concepts for Acoustic Cultural Heritages. Using Photogrammetric Models for the Preservation and Auralization of Historical Acoustical Spaces', in: *Proceedings of the 27th International Conference on Cultural Heritage and New Technologies, November 2022.* Heidelberg: Propylaeum.
DOI: xxxxxxx

Introduction

The photogrammetric or reconstructive modelling of cultural heritage sites is mostly focused on visually perceivable aspects, but if their intended purpose is the performance of cultural acts with a sonic emphasis, it is important to consider the preservation of the acoustic space of a cultural heritage for auralization purposes. This applies in particular to churches and theatres, which are popular objects for photogrammetric models. These models are suitable for the derivation of geometrical meshes and visual data, which can be used to automatically classify and map materials as well as surface properties for the purpose of integrating acoustical properties like absorption coefficients into an acoustic model. With the advancing conversion or destruction of historical acoustical spaces, it becomes even more important to preserve their sonic behaviour, as three-dimensional auralizations become widely applicable. This paper presents a concept for the processing of photogrammetric models and corresponding multimodal datasets for the preservation and auralization of acoustic cultural heritages.

Photogrammetric Aspects

The capturing of large and complex interior spaces using photogrammetric methods comes with difficulties and questions about how much time should be invested on which components. From an acoustical perspective, it is of most importance to properly capture the main room geometry including its ceiling, windows and columns. As the adjacent rooms might act as resonant spaces, especially during the appearance of high-energy emissions like the sound of low organ pipes, it is only necessary to capture their basic room geometries.

As the interior of historical cultural buildings are often based on predefined building principles and structures, repetitive elements can be captured once and replicated in the 3D model. For example, the capturing of all columns, benches, windows or wood panelling in a cathedral can be substituted by the more detailed capturing of one of these objects per class, as long as they appear repetitively in comparable shapes, materials and dimensions, while minor deviations are negligible. Another fundamental aspect for acoustical simulations is the surface structure of an object material, which can be derived from generalized models such as different types of brick, plaster or wood for flat surfaces. But as soon as surface structures or ornaments might have significant effects on the reflective behaviour of a material, it becomes important to capture their geometries more detailed as well, especially for repetitive occurrences.

Fig. 1. Photogrammetric 3D model of the interior of a church in Leipzig with different object classifications, a) repetitive objects (chairs & benches); b) different surface structures (painting, plastic and carpet); c) special geometrical features.

Room Impulse Response Measurement

The measurement of the room impulse response (RIR) can be done with simple equipment and is an inevitable component for acoustical simulations and auralizations, as it allows the analysis of the frequency-dependent wave propagation in a room. It describes the change of an emitted known signal until it arrives at a receiver, which can be achieved by a simple loudspeaker (= source) and a microphone (= receiver), but also by spherical loudspeakers and microphones in a more complex setup. The photogrammetric model can be used to determine the exact position of the source and the receiver for more comprehensible acoustical data. At this point it is necessary to use a laser range finder and protocol the spatial position as well as the angle of the loudspeakers and microphones that have been used including their properties. All acquired data shall be linked to each other, e.g. the source and receiver positions to the corresponding audio file, while as much information as possible shall be documented such as the interface, preamplifier, gain levels or noise reduction algorithms.

Classification of Materials and Textures

The highest accuracy in classifying materials and objects based on images can be achieved through artificial neural networks and machine learning approaches, which became widely available for different purposes. As the generation of photogrammetric models is based on photographs with a consistent quality and the assignment of their shooting positions in a three-dimensional space, it is possible to use an image classification algorithm on them and to extract the positional information for the generation of the acoustical model. The classified materials and textures can automatically be linked to acoustical properties, while the unknown materials properties can be edited manually, as the positional information of detected objects remain available. This pattern detection approach also allows segmentations, for example, to separate the wooden parts of a window from the glass surfaces, which makes the acoustical model more precise.

Fig. 2. Example for the automated classification of wooden parts and the texture of a window object as applied on a photogrammetric 3D model of the interior of a church in Leipzig. The errors around the frame can be ignored, as they are outside the borders of the window. The classification system will be applied to every segmented area within an object.

Model Refinement and Material Mapping

Acoustical models based on finite-element-methods (FEM) or raytracing are computationally intensive and need a room representation with as few geometrical faces as possible, while the texture information can be excluded for the mesh generation. To generate a proper mesh from the photogrammetric model for the acoustical modelling, it is necessary to remove any unwanted deformations of flat surfaces or irregularities, fill all holes appearing in the mesh, replace the rough repetitious objects with the detailed ones, remove suddenly appearing extremely small faces with disruptive angles and decimate the mesh. Recently developed photogrammetric methods based on deep learning like Neural Radiance Fields[1] might significantly reduce the manual effort of this process, especially regarding small interfering flaws in a mesh. The resulting model shall be imported into the simulation software, checked for errors or computational complexity and refined if necessary.
The classified materials with their corresponding positional information and absorption coefficients can be mapped onto the refined model in compliance with the protocol of the specific simulation software. At this stage of development, only COMSOL and ODEON are supported by the provided conversion tool.

Acoustical Simulations

After the mesh and the mapping of materials and textures have been derived from the photogrammetric 3D model, it becomes possible to analyze and simulate the acoustical behaviour of the spatial cultural heritage with the acquired room impulse responses. This becomes especially important for the reproduction of vocal or instrumental sounds in three-dimensional virtual environments, as the specific position of a person in a complex virtual building might have a high impact on the perceived

---

[1] Mildenhall, B. et al. (2022). 'NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis'. *Communications of the ACM*, 65(1), pp. 99-106, doi: 10.1145/3503250 (Accessed 15 August 2022).

sound. The analytical results can be implemented into an acoustical model, which provides an authentic auralization of a cultural heritage site and therefore higher immersive qualities. Furthermore, the data can be used for the study of musical instruments and their acoustic behaviour, especially with a focus on the organ, which is an instrument that is intended to be coupled to the building. The proposed concept will be used in the project *MODAVIS*[2] and its Framework, which is addressing the audiovisual virtualization of historical pipe organs.

Data Management and Model Implementation

To properly link the generated multimodal dataset, consisting of the original photogrammetric model, its photographs, the room impulse responses, source and receiver positions, the refined mesh as well as positional and acoustical data, it is important to provide a persistent access to properly defined data structures. For this reason, the standard for *Virtual Acoustic Objects* (VAO) has been developed, which is intended to concentrate multimodal data of acoustical objects into one file with consistent data structures, which allows to process musical instruments as well as acoustical spaces. The VAOs include the derived acoustical simulation data and can be decoded as playable virtual instruments or environmental auralizations of incoming sounds. In addition, it is possible to encode and link object information through persistent identifiers like related persons, properties, literature and more. The data management concept for virtual acoustic cultural heritages ensures the correct encoding of the generated data into a VAO and therefore its auralization, which can both be done with the developed tools in a code-based or interactive environment.

Musical Instruments and Cultural Heritage Sites in Metaverse Environments

With the emerging technologies that allow the capturing and auralization of three-dimensional wavefields it becomes a challenging task to create authentic virtual copies of musical instruments, as any virtual interaction should trigger a specific auditive reaction, starting with noises induced by simply touching the object or the sound of an intensively played note, which both can be highly complex. To implement these virtual instruments into multimodal virtual environments like the Metaverse, which is a central part of research in the field of Digital Organology, the VAO-Framework becomes relevant as an interface due to its focus on extensive export possibilities, which can directly be used by developers working with Unity or different engines that are specifically designed for AR-/VR-Applications. But the aspect of authenticity also affects the visual reproduction and audible behaviour of cultural heritage sites, as a perceivable divergence between the geometries of a virtual space and the behaviour of sound in this space might lead to a room-divergence effect[3], which confuses the psychoacoustic localization mechanism of a person in a virtual space and therefore disturbs the immersivity of a situation in the Metaverse. This does not only concern musical instruments and voices, but also any interaction with objects in a virtual environment, what emphasizes the relevance of a correct modelling of the acoustical behaviour of a cultural heritage site and therefore one of the main tasks of the framework. With a sufficient congruence between the visual and auditive model in

[2] Ukolov, D. (2022). *Project MODAVIS*. Available at: https://modavis.org/ (Accessed 18 July 2022).

[3] Werner, S. et al. (2016). 'A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events', *8th International Conference on Quality of Multimedia Experience (QoMEX)*, doi: 10.1109/QoMEX.2016.7498973 (Accessed 15 August 2022).

the Metaverse, the availability of an extending collection of interactive virtual instruments as well as historical spaces and their unique acoustical behaviours have a high potential to consolidate as essential cultural acts in the Metaverse. Furthermore, the omnipresent MIDI-Protocol enables an easy integration of widely available playing devices with haptic feedback like keyboards, guitars or flutes, which are able to emit silent signals that are being processed by the virtual instrument along with the positional information of the player that is situated in the Metaverse.

Use Cases and Further Research

Besides the preservation and study of the acoustical behaviour of cultural heritage sites, the multimodal digitization allows the integration and auralization of interactive VAOs in the Metaverse, be it in Augmented Reality (AR), Virtual Reality (VR), Cross Reality (XR) or in auditory-only virtual environments. One of many use cases intended by project MODAVIS is the development of an App that lets the user hear the (mostly silent) organ when entering a particular church, which can be realised through the auralization of a three-dimensional sound field, as the perceived sound field changes while the listener is moving inside the building. Another use case by this concept is to reproduce the acoustical space while playing the corresponding virtual organ with an AR-Overlay; additionally, a VR-Implementation of the VAO with head-tracking shall be developed.

Furthermore, the multimodal data and reproduction of the acoustical space will be used for the implementation of synthesized audio data into the room model, e.g. for the correct reflection behaviour of reconstructed sounds of corrupted or missing organ pipes. The simulation data shall be used for the analysis and reproduction of acoustical phenomena like fluctuations and resonance effects.

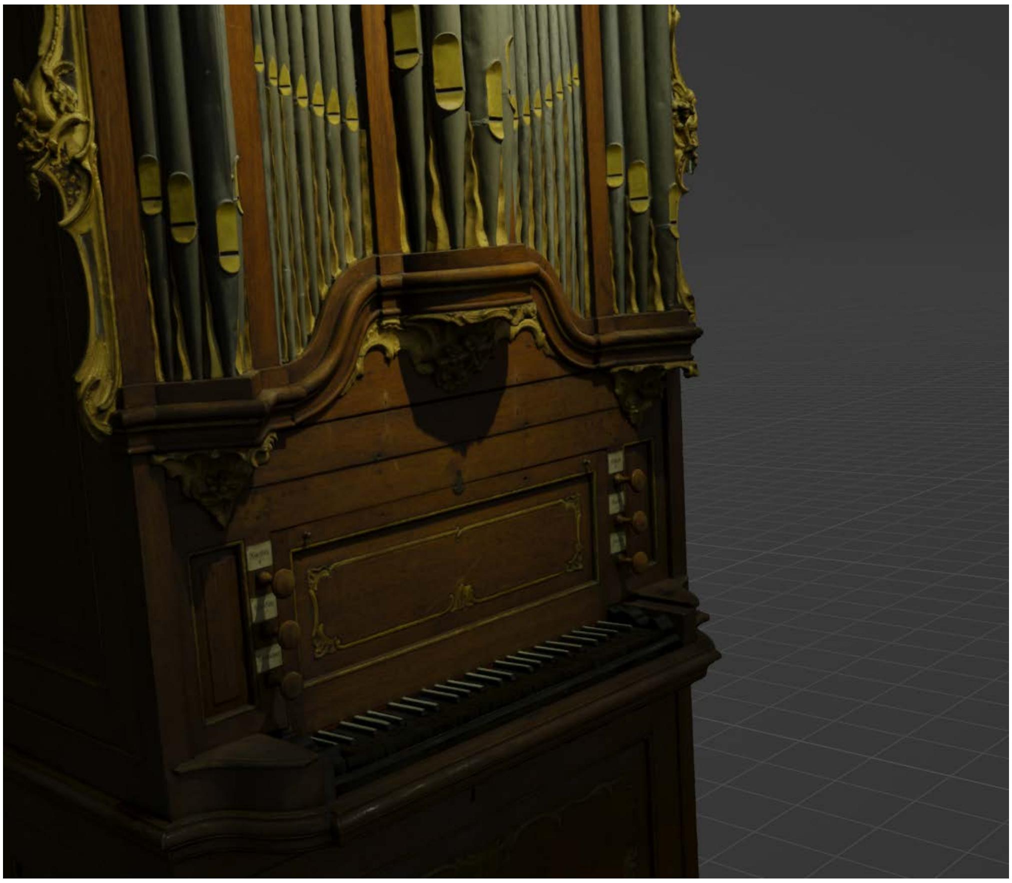

Fig. 3. Photogrammetric 3D model of an organ positive from the Museum for Musical Instruments of Leipzig University. Every single sound has been digitized, encoded into a VAO and can be implemented into the acoustical model of a historical acoustical cultural heritage site. This model is can also be used for an overlay in AR.

Conflict of Interests Disclosure

The author has neither financial nor personal conflicts to declare.

Author Contributions

Conceptualization: Dominik Ukolov
Data curation: Dominik Ukolov
Formal Analysis: Dominik Ukolov
Funding acquisition: <contributor names>
Investigation: <contributor names>
Methodology: Dominik Ukolov
Project Administration: Dominik Ukolov
Resources: Research Center for Digital Organology of Leipzig University
Software: Metashape, COMSOL, ODEON
Supervision: Prof. Dr. Josef Focht, Prof. Dr. Rolf Bader
Validation: <contributor names>
Visualization: Dominik Ukolov
Writing – original draft: Dominik Ukolov
Writing – review & editing: Dominik Ukolov

References

[1] Mildenhall, B. et al. (2022). 'NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis'. Communications of the ACM, 65(1), pp. 99-106, doi: 10.1145/3503250 (Accessed 15 August 2022)

[2] Ukolov, D. (2022). *Project MODAVIS*. Available at: https://modavis.org/ (Accessed 18 July 2022)

[3] Werner, S. et al. (2016). 'A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events', 8th International Conference on Quality of Multimedia Experience (QoMEX), doi: 10.1109/QoMEX.2016.7498973 (Accessed 15 August 2022)