arXiv:2302.05725v1 [cs.SD] 11 Feb 2023

基于摄影测量的文化遗产场所可参数化声学建模与可听化

预印本

Dominik Ukolov研究组：数字管风琴学 莱比锡大学，Augustusplatz 10，04109 莱比锡（德国）
dominik.ukolov@uni-leipzig.de

2023年1月31日

摘要

文化遗产场所的摄影测量与复原建模大多侧重于视觉上可感知的方面；但如果其预期用途是开展以声音为重点的文化活动，那么考虑如何保留其声学行为、使其能够以真实的方式被听见，就十分重要。这尤其适用于作为摄影测量模型常见对象的宗教与音乐会环境；这些模型包含几何和纹理信息，可用于定位和分类与声学相关的表面属性。随着历史声学空间不断改造或遭到毁坏，保留其独特的声音特质变得愈发重要，而三维听觉化的应用也日益广泛。本研究介绍了一种利用摄影测量数据进行声学建模的新方法的现状，并提出一套参数化处理流程，未来将作为带有图形用户界面的开源软件向公众开放。

关键词 声学 · 摄影测量 · 文化遗产 · 可听化 · 虚拟化

1 引言

乐器虚拟化是数字管风琴学中的一个新兴领域，但迄今主要局限于通过采样技术或物理建模来采集和合成乐器的声音。然而，MODAVIS项目旨在研究一种方法，以多模态且科学一致的方式将乐器数字化并虚拟化为三维声学对象，同时制定一项新标准：虚拟声学对象 [Ukolov, 2023]。这项研究尤其针对管风琴——所有乐器中最复杂、体积最大的乐器，同时也是最容易受到气候变化、军事冲突和经济危机影响的乐器。此外，由于管风琴与历史建筑和具有代表性的建筑物存在声学耦合，要全面理解其声音，就必须考虑周围环境；这些建筑物必须被视为共振体并进行精准建模。由于声音在建筑物内部传播时会受到其尺寸、材料和形状的影响，因此，摄影测量模型在生成用于可听化的交互式虚拟声学对象和声学房间模型方面前景广阔。为解决这一问题，本文提出的方法以分类掩膜与其三维对应物之间的点投影为基础，并为其标注声学属性；整个过程通过一个参数化流程完成（见图 1）。

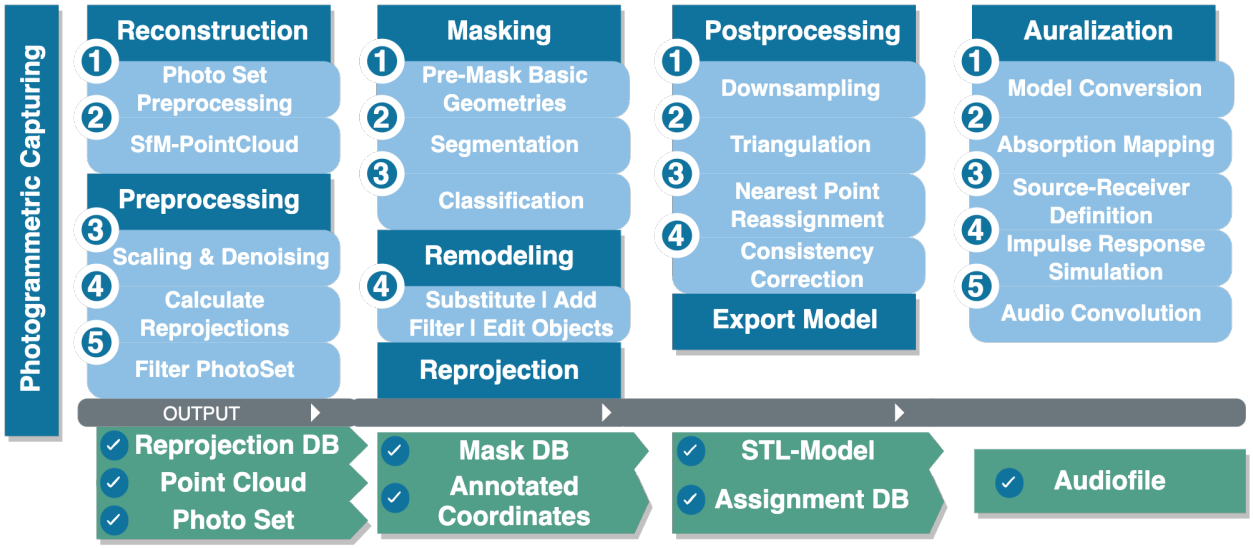


图1：该处理流程包含多个处理步骤，可通过各种参数进行修改；程序生成的输出数据还可进一步分析、编辑，或用于其他实例。模块化结构支持优化现有步骤并添加更多步骤，这些步骤将在 MODAVIS 项目期间进行评估和实施。

2 相关工作

Llorca-Bofí 等人开展了关于使用摄影测量模型进行室内声学仿真的相关研究 [2022], , 他们采用基于 Siltanen 等人研究 [2008] 的几何简化方法，并使用 RAVEN 插件评估了多种建模方案和材料分配 [Schröder 和 Vorländer, 2011], ；而可听化的基础理论则由 Vorländer 进行了研究 [2020]。Chen 等人利用 SoundSpaces 2.0 [2022], , 在网格数据集上通过双向路径追踪连续渲染脉冲响应，并在对 Replica 数据集中的一个物体进行声学测量后评估了这些响应 [Straub 等人，2019]。然而，下文所述的方法与这些研究不同：它基于通过点重投影进行重建，并利用神经网络进行多重分割；据目前所知，这是首个采用这些方法生成声学模型的方案。

3 摄影测量声学建模工具包

为了使处理流程能够通过参数进行配置并提供实时视觉预览，同时让不同专业水平的用户都能轻松操作声学建模过程，我们开发了摄影测量声学建模工具包（PAMT）（见图2），该工具包支持手动和半自动处理步骤。

3.1 预处理

PAMT 使用 COLMAP [Schönberger 和 Frahm， 2016] 根据一组照片计算点云并提取相机参数。创建新项目文件后，首先会利用这些数据生成点重投影数据库，其中保存了所有处理流程设置和单项建模操作。接下来，可以根据标记、测量数据或模型对齐结果对提取的数据进行缩放，然后根据照片之间的重叠情况筛选照片集，以减少后续处理所需的照片数量。

3.2 对象分割

精简后的照片集可在图形用户界面中用于手动或自动分割对象。手动分割通过多边形遮罩完成，并可使用 Canny 算法调整边缘检测 [Xu 等人，2017]。自动分割采用 Mask R-CNN [He 等人， 2017], , 并使用高于可调阈值的输出生成遮罩。设置遮罩后，可手动指定或根据分类建议为其分配对象类别、对象和材料；之后，可以合并或拆分多个遮罩，因为所有遮罩始终都可编辑。处理完一张照片后，遮罩中的点将被扩展到其他

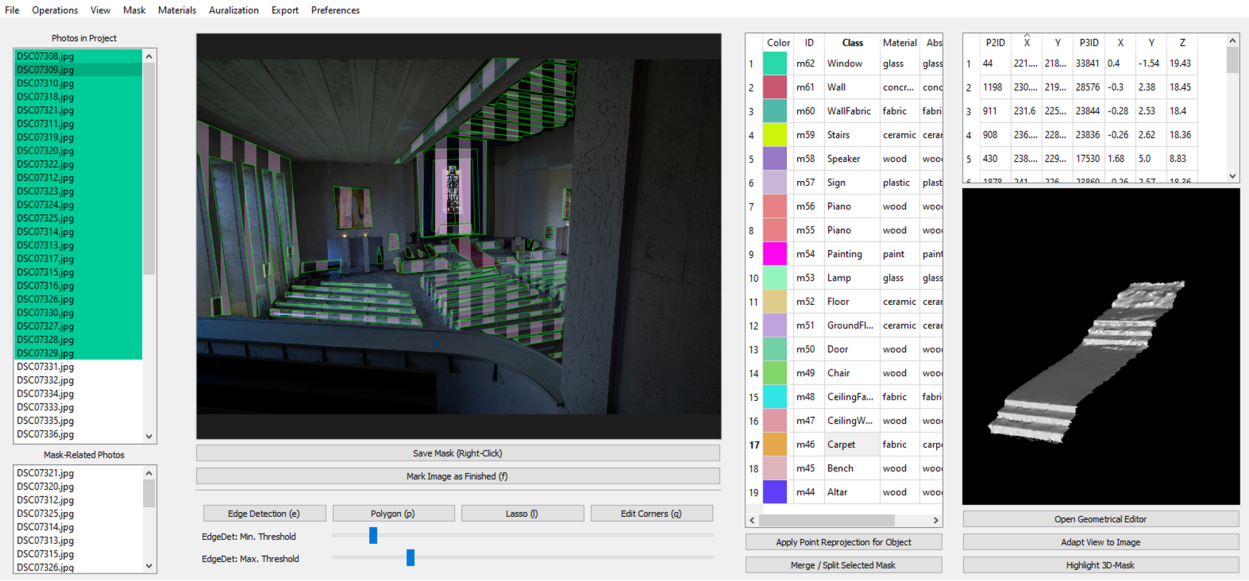


图2：PAMT的主界面包括用于摄影测量计算的精简照片集（左上）、遮罩窗口及其控件（左中），以及包含遮罩区域、分类对象、材料和吸声系数的列表（右中）。应用重投影后，可以在左下和右上查看相关图像以及对应的局部点和全局点；这些点将在图像和下方的三维查看器中标记出来。该查看器具有交互功能，可显示所选遮罩及其重投影对应的对象；而“几何编辑器”会打开一个外部窗口，用于对象修改、平面分割及其他几何操作。在此截图中，从遮罩列表中选中了地毯，并应用了相应的重投影，从而将关联的点坐标渲染为当前遮罩进度的预览。

照片：将靠近重新投影后的二维像素、且位于遮罩内的三维点标识投影至这些照片中。由此可为全局表面预先生成局部遮罩，大幅减少手动操作所需的时间。可在设置中选择、下载和更新用于自动检测与分割的预训练模型；工具包集成了 MMLab [Chen 等人， 2019]，以确保能够实现针对此任务的最先进模型。

3.3 材料分类与吸声系数

设置掩膜后，可以手动定义材料，也可以分两个阶段采用神经网络进行分类；目前使用的是基于 ResNet-18 [He et al., 2016]架构的分类器，该架构分别在用于识别材料的 MINC 数据集[Bell et al., 2015] 和用于识别纤维状、大理石纹或多孔等纹理属性的 DTD 数据集 [Cimpoi et al., 2014]上进行训练。为了给具有纹理的材料指定频率相关吸声系数，我们处理了一个包含 2573 条记录的专用数据库 [德国联邦物理技术研究院] ，以便用于推荐可能合适的测量值；这可以通过自然语言处理（NLP）实现，方法是根据向量表示计算输入字符串与声学测量材料名称之间的余弦相似度。

3.4 点重投影

重投影背后的基本方法是将局部二维坐标分配给全局三维点标识，以便利用这些标识映射照片中的局部点。多边形遮罩定义了一组坐标，由此可以推导出相应的标识及其分配的二维坐标。因此，一组照片的局部坐标始终与 PAMT 中的全局坐标相对应，反之亦然。为此开发了数据库访问结构，只需提供任意几何输入数据，即可检索和处理目标数据。

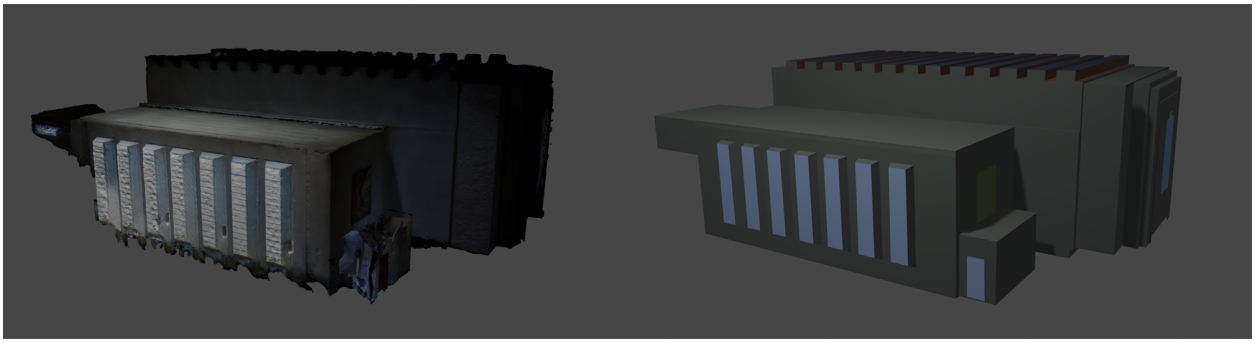


图 3：采用所提出的处理流程，可以将摄影测量数据得到的点云重新建模为几何简化后的点云（右），经过一致性校正后即可用于声学模拟。教堂内部的摄影测量模型（左）从外部视角拍摄，拍摄时照明条件较差，且使用了低质量相机，因此部分区域存在误差。尤其明显的是，平面分割和边界计算会导致形状发生显著变化（见两个模型的左侧和右下侧）；这些变化可能像本例一样带来预期的结果，但仍需在不同的低质量和高质量数据集上进一步评估。右图中可见的材料——玻璃（浅蓝色）、木材（棕色）、织物（深蓝色）和绘画（黄色）——均被赋予处理流程中定义的表面；建筑物外部未被拍摄，在当前开发阶段也与研究无关。

3.5 对象修改

由于成排椅子等物体常常重复出现，且并非每处都能被详细采集，因此可以通过指定物体，或导入另一个点云或网格，用细节更丰富的物体替换点云中不准确的物体外观。可以通过估算物体相对于房间的位置或手动放置来确保其正确定位；此外，还可以筛选或添加对象类别或单个物体，从而实现不同房间情境的声学模拟。

3.6 点云与表面重建

通过遮罩定义对象坐标后，可以使用 Blender API 生成标注点，并在图形用户界面中通过交互式控件进行渲染，从而显示对象类别或单独的遮罩区域。点云去噪采用基于评分的方法 [Luo and Hu, 2021], ，而渲染和参数化过程则使用 Open3D [Zhou 等人, 2018], ，包括降采样和三角剖分，目前采用球旋转 [Bernardini 等人, 1999] 或泊松 [Kazhdan 等人, 2006] 算法。平面分割是模型构建的基础（见图 3），使用 RANSAC 算法 [Fischler and Bolles, 1981], 完成；不过，由于点集在这些过程中发生了变化，最终会将最近的点分配给相应的标识。最后，将模型导出为 STL 格式并进行验证，以防止可听化过程中出现孔洞或体积不当等不一致问题。此外，还将生成一个参照数据库，用于指向 STL 文件中每个元素的声学表面属性。

3.7 声学仿真与可听化接口

要从特定听者位置仿真房间声学情境，可通过 Pyroomacoustics [Scheibler 等人, 2018], 的接口实现带注释模型。该接口会将编码后的几何数据和声学数据转换为兼容的数据结构。通过定义声源和接收器的坐标及其发射属性，设置任意的声源位置和接收器位置后，即可应用镜像声源法 [Allen and Berkley, 1979] 和声学射线追踪 [Krokstad 等人, 1968]来仿真脉冲响应，并将其与输入音频信号进行卷积。利用摄影测量计算，还可以从照片所捕捉的精确视角进行听觉化，因为相应的相机位置和朝向均已知。

4 未来工作

计划使用有效的声学测试数据集评估 PAMT 的性能，并在 auralization 引擎中加入散射系数和指向性模式等重要的心理声学和声学参数；在当前开发阶段，这些参数通常被视为次要因素。

几何处理的配置和方法总体上将进行优化，以实现尽可能低的错误率。在 MODAVIS项目中，计划为宗教环境和管风琴特定部件创建对象与材质分割数据集；此外，还将研究能否通过物理建模模拟并泛化不同管风琴音管类型的指向性模式。工具包还将支持使用 RGB-D 数据和非摄影测量生成的三维模型，并借助虚拟图像渲染呈现这些模型；同时，也将能够在正射地图上标示视角，以避免对重复结构产生混淆。后续阶段将考虑发布用于标注和可听化的 Blender 模块及 Unity 接口。完成评估和重大改进后，PAMT 将在 Git Hub 上发布 [Ukolov，2022]。

参考文献

J. B. Allen 和 D. A. Berkley。用于高效模拟小房间声学的镜像法。美国声学学会期刊，65(4):943–950，1979。S. Bell、P. Upchurch、N. Snavely 和 K. Bala。利用材料上下文数据库，在真实场景中识别材质。载于*IEEE*计算机视觉与模式识别会议论文集，第3479–3487页，2015。F. Bernardini、J. Mittleman、H. Rushmeier、C. Silva 和 G. Taubin。用于表面重建的球旋转算法。*IEEE*可视化与计算机图形学汇刊，5(4):349–359，1999。C. Chen、C. Schissler、S. Garg、P. Kobernik、A. Clegg、P. Calamia、D. Batra、P. W. Robinson 和 K. Grauman。SoundSpaces 2.0：用于视听觉学习的仿真平台。*arXiv:2206.08312*，2022。K. Chen、J. Wang、J. Pang、Y. Cao、Y. Xiong、X. Li、S. Sun、W. Feng、Z. Liu、J. Xu 等。MMDetection：Open MMLab 检测工具箱与基准。*arXiv*预印本 *arXiv:1906.07155*，2019。M. Cimpoi、S. Maji、I. Kokkinos、S. Mohamed 和 A. Vedaldi。描述真实场景中的纹理。载于*IEEE*计算机视觉与模式识别会议论文集，第3606–3613页，2014。M. A. Fischler 和 R. C. Bolles。随机采样一致性：一种模型拟合范式及其在图像分析和自动制图中的应用。*ACM*通讯，24(6):381–395，1981。K. He、X. Zhang、S. Ren 和 J. Sun。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集，第770–778页，2016。K. He、G. Gkioxari、P. Dollár 和 R. Girshick。Mask R-CNN。载于*IEEE*国际计算机视觉会议论文集，第2961–2969页，2017。M. Kazhdan、M. Bolitho 和 H. Hoppe。泊松表面重建。载于 A. Sheffer 和 K. Polthier 编，《几何处理研讨会》。欧洲图形学协会，2006。doi:10.2312/SGP/SGP06/061-070。A. Krokstad、S. Strom 和 S. Sørsdal。利用光线追踪技术计算房间声学响应。声音与振动期刊，8(1):118–125，1968。J. Llorca-Bofí、J. Heck 和 M. Vorländer。用于可听化的三维摄影测量——几何简化与材质分类方法。第1–7页，2022年9月。URL https://www.researchgate.net/publication/ 367026449。S. Luo 和 W. Hu。基于分数的点云去噪。载于*IEEE/CVF*国际计算机视觉会议论文集，第4583–4592页，2021。德国联邦物理技术研究院。房间声学吸声系数数据库。URL https://www.ptb.de/cms/ptb/fachabteilungen/abt1/fb-16/ag-163/absorption-coefficient-database.html。访问日期：2023-01-31。R. Scheibler、E. Bezzam 和 I. Dokmani´c。Pyroomacoustics：用于音频房间模拟和阵列处理算法的 Python 软件包。载于*2018*年*IEEE*国际声学、语音与信号处理会议（*ICASSP*），第351–355页。IEEE，2018。D. Schröder 和 M. Vorländer。RAVEN：用于交互式虚拟环境可听化的实时框架。载于*Forum acusticum*，第1541–1546页。丹麦奥尔堡，2011。J. L. Schönberger 和 J.-M. Frahm。重新审视运动恢复结构。载于*IEEE*计算机视觉与模式识别会议论文集，第4104–4113页，2016。S. Siltanen、T. Lokki、L. Savioja 和 C. Lynge Christensen。房间声学建模中的几何简化。*ActaAcustica unitedwith Acustica*，94(3):410–418，2008。J. Straub、T. Whelan、L. Ma、Y. Chen、E. Wijmans、S. Green、J. J. Engel、R. Mur-Artal、C. Ren、S. Verma 等。Replica 数据集：室内空间的数字复制体。*arXiv:1906.05797*，2019。

D. Ukolov。MODAVIS，2022。URL https://github.com/modavis-project/。访问日期：2023-01-31。
D. Ukolov。虚拟声学对象，2023。URL https://modavis.org/virtual-acoustic-objects。访问日期：
2023-01-31。M. Vorländer。可听化：声学基础、建模、仿真、算法与声学虚拟现实。Springer Nature，
2020。Z. Xu、X. Baojie 和 W. Guoxin。基于 OpenCV 的 Canny 边缘检测。载于*2017*年第*13*届*IEEE*国际电子测量与仪器会议（*ICEMI*），第53–56页。IEEE，2017。Q.-Y. Zhou、J. Park 和 V. Koltun。
Open3D：用于三维数据处理的现代库。*arXiv:1801.09847*，2018。