Three-dimensional point-cloud room model for room acoustics simulations

Markovic, Milos; Olesen, Søren Krarup; Hammershøi, Dorte

Published in:
Meetings on Acoustics. Proceedings

DOI (link to publication from Publisher):
10.1121/1.4800237

Publication date:
2013

Document Version
Early version, also known as pre-print

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Architectural Acoustics

Session 4pAAa: Room Acoustics Computer Simulation II

4pAAa12. Three-dimensional point-cloud room model for room acoustics simulations

Milos Markovic*, Søren K. Olesen and Dorte Hammershoi

***Corresponding author's address: Aalborg University, Aalborg, 9000, Denmark, Denmark, mio@es.aau.dk**

Telepresence applications require communication with the feeling of being together and sharing the same environment. One important task in these applications is to render the acoustics of the distant room for the telepresence system user. This paper presents a fast method for the room geometry acquisition and its representation with a 3D point-cloud model, as well as utilization of such a model for the room acoustics simulations. A room is scanned with a commercially available input device (Kinect for Xbox360) in two different ways; the first one involves the device placed in the middle of the room and rotated around the vertical axis while for the second one the device is moved within the room. Benefits of both approaches were analyzed. The device's depth sensor provides a set of points in a three-dimensional coordinate system which represents scanned surfaces of the room interior. These data are used to build a 3D point-cloud model of the room. Several models are created to meet requirements of different room acoustics simulation algorithms: plane fitting and uniform voxel grid for geometric methods and triangulation mesh for the numerical methods. Advantages of the proposed method over the traditional approaches are discussed.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800237]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Users' expectations of a present-day communication technology go beyond services that give a possibility of long distance, real time conversation. Communication with the feeling of being together and sharing the same environment is desired [1]. The EU project BEAMING [2] is currently addressing the issue of improving immersive communication interfaces. BEAMING is a collaborate research project where the goal is to give people (Visitors) a real sense of physically being in a remote location with other people (Locals) without physically travelling. Simultaneous streams of data from the destination site to the Visitor's perceptual apparatus, and from the actions and state of the Visitor to the destination site, cohere together to form a unified virtual environment representing the physical space of the destination in real-time, a destination that now includes the beamed people. Therefore, a concept of Visitor presence at the destination is to be achieved by the substitution of real sensory data with virtually generated sensory data. The substitution is "successful" to the extent that the Visitor forms percepts from the rendered sensory data and responds to and acts upon these as if they were real [3].

Spatial sound plays an important role if a concept of presence at the remote location is desired. Different techniques are used for spatial sound rendering [4]. Most of them are based on the same principle: modelling of the sound field and reproducing the rendered sound. Modelling relies on knowledge of sound propagation behavior in the acoustical space – room acoustics simulation, while rendered sound reproduction involves binaural cues of the human hearing [5]. Fast room geometry acquisition suitable for the real-time room acoustic simulations is required when dynamic, interactive environments are to be acoustically rendered.

ROOM ACOUSTICS MODELLING

Computer modelling of room acoustics was proposed in 1960's by Schroeder [6], and has since been used for a variety of virtual acoustics applications [7]. Virtual acoustics can be created with perceptually or physically based modelling techniques. Perceptually based virtual acoustic rendering is often performed without any knowledge of the room geometry, and perceptually optimized efficient algorithms are used. In other words, perceptually based virtual acoustics endeavours to reproduce only the perceptually salient characteristics of reverberation [8]. Although the physically based approach also utilizes knowledge of the human sound perception, it seeks to simulate sound propagation from the source to the listener accurately for a given room.

Two main approaches in the physically based room acoustics modelling are wave-based and geometrical acoustics [9]. The first one is more accurate and it is based on solving the actual wave equation numerically. This approach is computationally demanding and the workload grows rapidly as a function of the frequency. Usually, it is used in non-real time applications to simulate low frequency sound fields [7]. For the wave-based techniques, either the space or its bounding surfaces are discretized to small elements and the interaction between them is modelled. These techniques are often called element-based methods and the most used are: finite element method (FEM), boundary element method (BEM) and finite-difference time domain methods (FDTD) [10,11]. Using geometrical acoustics sound waves are considered as rays or lines and the wavelength is either neglected or assumed high enough for specular reflections really to take place. Thus, some low-frequency phenomena are missing and need to be simulated in addition. Examples of the geometrical acoustics techniques are image source [12], ray-tracing [9] and beam-tracing [13] methods.

All of the physically based room acoustics modelling methods require the room geometry to be parametrically described prior to a sound transmission calculation. This is a highly room-specific task and rather time consuming if a complex geometry is to be described. Here, a generic method for an arbitrary room geometry acquisition is presented. The method exploits a depth sensor of a Kinect device that provides a point based information of the scanned room interior.

POINT-CLOUD

The whole procedure from a room scanning to the sound transmission module of a room acoustics modelling software consists of two steps. First, the room interior is scanned with depth-camera (Kinect) using 3D scanning software. Then, the 3D point cloud of the room is processed using the Point Cloud Library (PCL) [14], in order to make an optimal room geometrical model for the sound transmission calculation, Figure 1.

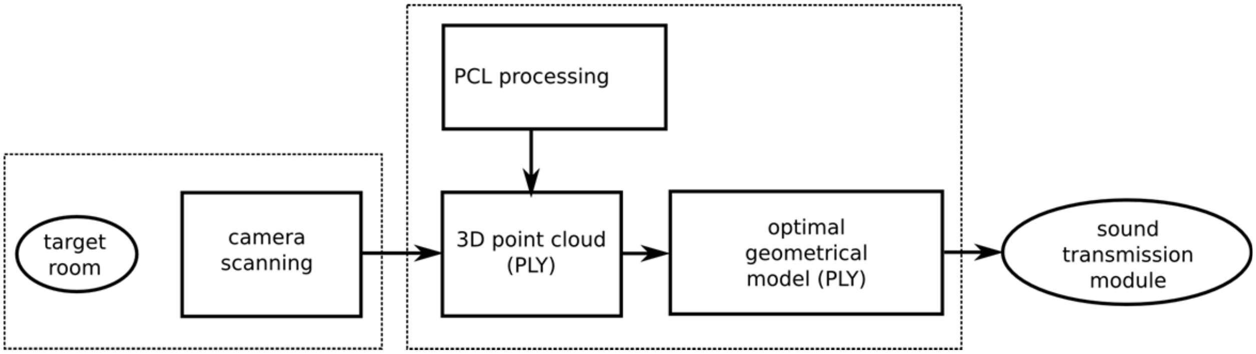

FIGURE 1. A 3D point-cloud room model obtaining procedure

The proposed method is independent of the input device ("depth camera"). In general, all devices that provide a boundary description of a scanned room interior can be used to obtain a 3D point-cloud model. Still, there are several commercially available types of cameras that employ different techniques. Although all techniques provide a point based information of a scanned room interior, several camera parameters have to be taken into account.

Depth-camera technologies

Cameras that can acquire a continuous stream of depth images are nowadays commonly available. They can be, to some degree, a good substitution for the expensive 3D laser scanners when a spatial geometry of the existing room is to be captured. In addition, most of them provide RGB information of the scanned surfaces which can be used for the visual rendering. That allows simultaneous rendering of the visual scene and the geometry acquisition needed for the acoustical simulation. Also it opens a possibility of the visual data utilization for recognition of the scanned surface material and thus acquisition of its acoustical properties [15].

Selection of the most suitable depth camera is based on the several parameters: the employed technology, colour image and depth-map resolution, level of accuracy, depth measurements density and the price of the camera. Tree different camera technologies are considered: stereo vision exploited in Bumblebee XB3 camera (SV, three RGB cameras), structured light used in Kinect for Xbox360 (SL, one RGB sensor, one infra-red projector, one infra-red sensor) and time of flight employed in PMD CamCube (ToF, using light pulses). All of them provide geometrical description of the boundaries while the visual output is different: SV provides stereo pair of RGB images, SL provides RGB image while ToF provides grayscale intensity image [16, 17]. Additional properties are given in Table 1.

TABLE 1. Properties of three considered cameras

	Max. resolution [pixels]	**Max. frame rate [fps]**	**Working range [m]**	**Depth measurements**	**Price range [approx. $]**
Bumblebee XB3	1280x960	15	0.5 – 4.5	Inferred from color information	2000
Kinect for Xbox360	640x480	30	0.6 - 5.0	Depth frames registered to RGB frames	200
PMD CamCube	200x200	25	0.3 - 7.5	Directly sampled from the scene	10000

The Kinect device represents a good choice for room geometry acquisition. It can densely cover the full scene with its infra-red structured light pattern and thus provides depth measurements with the high resolution (3mm point-to-point distance). The obtained homogeneous depth map is associated with the colour frame which allows merging the visual and acoustical rendering. The Kinect is used as a standard device in the BEAMING project.

Room scanning

A room is scanned with Kinect in two different ways: first one involves the device placed in a middle of the room and rotated around the vertical axis while for the second one the device is moved within the room. Both approaches are analysed in terms of post-processing potential, accuracy, required scanning time and size of the obtained data. First approach is more generic and it's applicable for any room geometry. There's no need for a human operator; the device can be mounted on the automatic rotating platform. The whole scanning procedure takes approximately two minutes. A main drawback of this scanning method is the existence of shadows. Areas which are shadowed by other surfaces and not visible from the camera standing point are not scanned and thus, the model contains “holes”. Also, due to the camera's angle of view, blind spots exist below and above the camera standing point leaving the model without data in these areas, Figure 2. This can be solved with additional processing of the obtained data by a surface (plane) recognition algorithm. When the planes are recognised a model can be supplemented with the artificial data that fit the existing planes.

FIGURE 2. 3D point-cloud room model obtained by a camera placed in centre of the room and rotated around the vertical axis to form a full hemisphere

An alternative is to have the camera move through the room and not just rotate. That will provide more details from different angles. There's no need for additional plane fitting, but a movable sensor requires scanned scenes to be registered in the same model which takes time and slows down the scanning procedure. Also, a human operator is needed to control the scanning and decide on the model quality by visual observation while the scanning takes place.

Data description and manipulation

The device's depth sensor provides a set of points in a three-dimensional coordinate system which represents scanned surfaces of the room interior. These points are stored in 3D point cloud data formats, well known in the field of computer graphics. The most used formats are PCD (Point Cloud Data) and PLY (Polygon File Format). They have a similar structure and consist of a header where a variety of model's properties at the level of points can be defined, e.g. (x,y,z) coordinates of the vertices, RGB triple of a point colour, a number of element vertex, (x,y,z) coordinates of the sensor position, a normal based on a certain number of neighbour points etc. Second part of the data format is a list where each row represents the values of the point's properties defined in the header. These values can be given as ASCII or a binary file formats. A PLY file obtained from the sensor output after scanning the room is shown in Table 2. Free software for 3D scanning called Scanect is used [15]. The resulting PLY contains 4998006 vertices represented by ASCII and defined with (x,y,z) coordinates. In addition, position of the sensor for each point is provided (nx,ny,nz), as well as its colour (RGB). Position information of the sensor allows it to be moved through the room while the data are registered on the same model.

TABLE 2. PLY file format

ply
format ascii 1.0 element vertex 4998006 property float x property float y property float z property float nx property float ny property float nz property uchar red property uchar green property uchar blue element face 3332004 property list uchar uint vertex_indices end_header -0.309623 2.33571 2.17339 -0.908879 -0.297005 -0.292794 141 128 110 -0.306158 2.34149 2.17819 -0.908879 -0.297005 -0.292794 141 128 110 -0.300281 2.34501 2.17541 -0.908879 -0.297005 -0.292794 141 128 110 -0.297869 2.34276 2.16782 -0.908879 -0.297005 -0.292794 141 128 110 -0.301334 2.33699 2.16302 -0.908879 -0.297005 -0.292794 141 128 110 -0.307211 2.33347 2.1658 -0.908879 -0.297005 -0.292794 141 128 110 . . .

3D POINT-CLOUD MODEL OF A ROOM GEOMETRY

The Kinect output is post-processed in order to obtain a 3D point-cloud model of the room. Different post processing is applied for the optimal 3D point-cloud models to be used as an input for a sound transmission module of a different physically based room acoustics modelling methods. In general, granular representation of a room's inner surfaces makes it possible to have a high level of details when creating a geometry model. A high density of points makes it possible to recognize fine details and sharp edges of the surfaces as well as transitions between adjacent areas. Using a different resolution for a room model, it is possible to take into account any frequency-dependent geometry. Thus, on the basis of the original model, a downsampled point-cloud models can be created with less details, e.g. for low frequency simulations. Also, an advantage of a point-based geometry representation is an ability to store information about acoustical properties of each point directly in the 3D point-cloud model just as another property of the PLY header. Eventually, this can facilitate calculation of the sound transmission within the room.

Numerical methods

The point-cloud model can be efficiently used to create a triangulation mesh for the numerical acoustical methods. PCL provides a surface triangulation algorithm of a point cloud with normals to create a triangle mesh. It is based on projections of the local neighbourhoods. By maintaining a list of points from which the mesh can be grown and extending it until all possible points are connected [18], a concave hull can be created that represents room boundaries and the inner surfaces, Figure 3. The planar triangles that the hull is made of enable analytical integration when the acoustical boundary conditions are set. For other numerical methods, a whole 3D model including boundaries, inner surfaces but also the inner empty space is divided into small space units – voxels. Different type of voxels (with and without points of the point-cloud model) define a space grid that can be used in element based modelling, e.g. the finite element method.

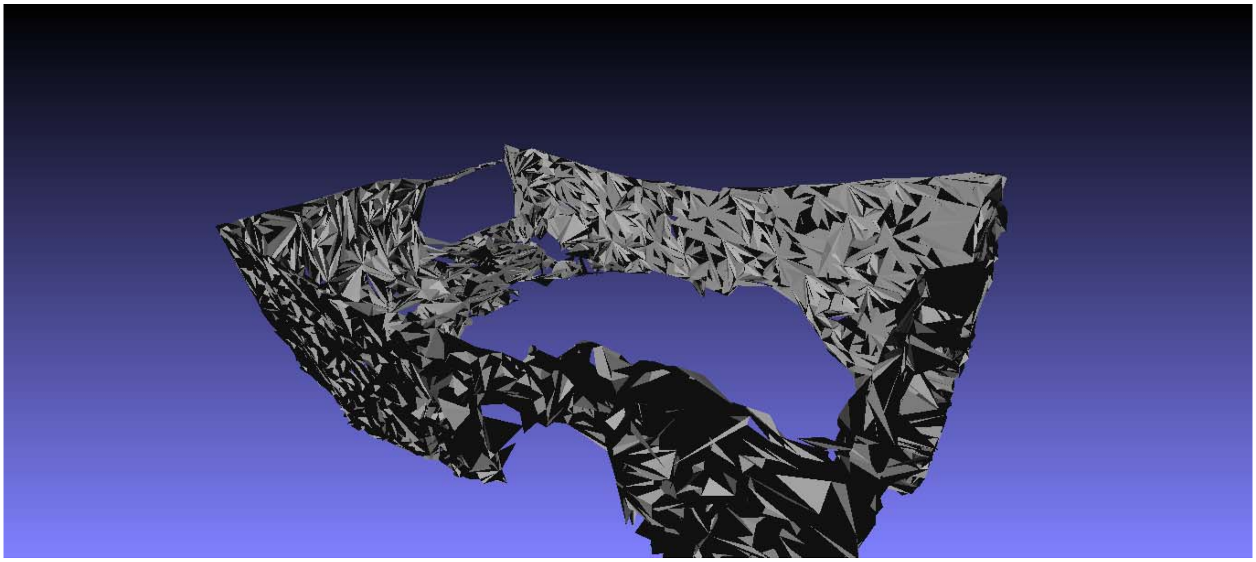

FIGURE 3. Triangular mesh representation of a room

Geometrical methods

Two approaches for geometrical simulations are presented. The first one is more conventional and involves plane recognition and surfaces' normal estimation. PCL is used in order to define important plains/surfaces for the acoustical simulation (walls, floor, ceiling...). Then, for all surfaces of interest, normals are estimated according to the defined neighborhood of a point where the normal is calculated. In this way, a set of normals from the same surface can be obtained only by defining different neighborhood size and thus considering the level of details. This can be useful when a frequency dependent simulation of a reflection is desired. A 3D model defined in this way is useful for standard image source or ray-tracing method.

The second approach strives to use the potential of a granular geometry representation in its original form for the discrete ray-tracing simulation algorithm. A voxel grid of a room is created and filled with the obtained point cloud from the depth-camera, Figure 4. Each voxel is defined as an occupied (contains point cloud data), or empty (doesn't contain point cloud data). Depending on the voxel type, different properties are assigned in order to define the behaviour of a sound ray when it reaches the voxel, e.g. reflection, diffraction, scattering, transmission etc. The voxel grid resolution defines the wanted level of details. A grid can be defined as a uniform grid when all space units are cubes of the same size but also as a hierarchical structure (k-d tree or octree). The latter one improves efficiency of the sound transmission module by speeding up the ray traversal algorithm.

DISCUSSION

By using point-cloud representation of an acoustic environment, it is possible to calculate its acoustics from the interior information obtained from the depth camera. This approach is suitable for the run-time applications since it doesn't include any method based on an audible excitation that can disturb running audio. Also, the point-cloud model can be updated in real time according to the scene changes detected by depth-camera. This allows efficient acoustical simulation of dynamic, interactive environments. Although only geometrical information of a room is provided, high amount of surfaces' details leaves possibility for implementation of material recognition algorithms that involve semantic mapping [15]. This can provide information of surfaces' reflective properties at the point level. Also, a high amount of details allows a good approximation of complex geometries (porous materials, rough surfaces etc.) and thus, more natural simulation of wave phenomena like diffraction and scattering.

ACKNOWLEDGMENTS

BEAMING is a collaborative project supported by the European Commission under the FP7 ICT Work Programme. The authors would like to acknowledge BEAMING collaborative partners from Dept. of Computer

Science, University College London and in particular Fabrizio Pece, PhD Student at this department, for his help in room geometry acquisition using depth cameras.

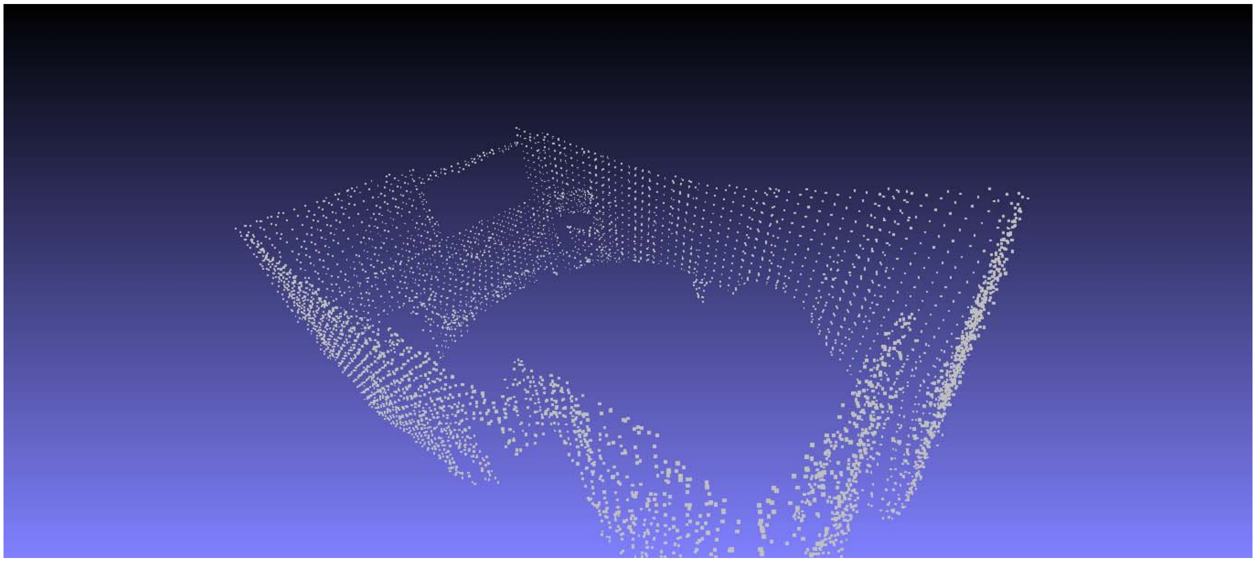

FIGURE 4. Uniform voxel grid representation of a room

REFERENCES

1. Y. A. Huang, J. Chen, J. Benesty, "Immersive audio schemes", IEEE Signal Processing Magazine, **28**, 20-32 (2011).
2. An official web site of the BEAMING project, http://beaming-eu.org/home.
3. M. Slater et al., "How We Experience Immersive Virtual Environment: the Concept of Presence and Its Measurement", Anuario De Psicología**,** **40**, 193-210 (2009).
4. D. R. Begault, "3D sound for virtual reality and multimedia", (AP Professional, Cambridge, MA, 2000).
5. H. Møller, "Fundamentals of Binaural Technology", Applied Acoustics, **36**, 128–171 (1992).
6. M.R. Schroeder, B.S. Atal, C. Bird, "Digital computers in room acoustics", Proceedings 4th ICA 62, Copenhagen, p. M21.
7. M. Vorländer, "Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality", (Springer, 2007).
8. W. G. Gardner, "Reverberation algorithms", in *Applications of Digital Signal Processing to Audio and Acoustics*, edited by M. Kahrs and K. Brandenburg, (Kluwer Academic Publishers, Norwell, MA, 1997), pp. 85–131.
9. L. Savioja, "Modeling Techniques for Virtual Acoustics", PhD thesis, (Helsinki University of Technology, report TML-A3, 1999).
10. R. D. Ciskowski, C.A. Brebbia, "Boundary Element Methods in Acoustics", (Springer, 1991).
11. D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems", J. Acoust. Soc. Am., **98**(6), 3302–3308 (1995).
12. J.B. Allen, D.A. Berkley, "Image method for efficiently simulating small-room acoustics", J. Acoust. Soc. Am., **65**(4), 943–950 (1979).
13. T. A. Funkhouser, et al., "A beam tracing method for interactive architectural acoustics", J. Acoust. Soc. Am., **115**(2), 739–756 (2004).
14. Point Cloud Library, http://pointclouds.org/.
15. D. Filiat, et al. "RGBD object recognition and visual texture classification for indoor semantic mapping", Proceedings of the 4th International Conference on Technologies for Practical Robot Applications, (Woburn, Massachusetts, USA, 2012).
16. D. Scharstein, R. Szeliski, "High-accuracy stereo depth maps using structured light", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (Madison, WI, USA, 2003).
17. C. Beder, et al. "A Comparison of PMD-Cameras and Stereo-Vision for the Task of Surface Reconstruction using Patchlets", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (Minneapolis, Minnesota, USA, 2007).
18. 3D scanning software, http://skanect.manctl.com/