Low-Cost Acoustic Sensor Array for Building Geometry Mapping using Echolocation for Real-Time Building Model Creation

T.A. Sevilla¹, W. Tian¹, Y.Fu²*, W. Zuo*⁺
¹ Department of Civil, Architectural, and Environmental Engineering
University of Miami, Coral Gables FL, 33146, United States
² Current Affiliation: University of Colorado Boulder
⁺Corresponding Author: Wangda.Zuo@colorado.edu

SUMMARY
Validation of result accuracy for indoor environment modeling and simulation usually requires high-quality data collected by expensive sensors and human operators. This becomes a problem when dealing with building retrofits, specifically when obtaining building geometry for later use in airflow, envelope, or human-building interaction simulations. Thus, we developed a low-cost acoustic sensor array (less than $70) to automatically detect and map building geometry using echolocation. Our focus of this research is to allow building modelers to obtain geometric as well as spatial information of existing buildings. The underlying hardware uses the Open Source Computer Vision Library which allows multi-core processing and enables hardware acceleration of the underlying heterogeneous compute platform, allowing us to perform distributive calculation among multiple devices. Likewise, this is useful when clustering devices for aggregate data collection for use in large mapping projects. Overall, this study proposes a prototype for future work in this field.

INTRODUCTION
Advances in architectural design have caused a rise in complex building geometries. Similarly, the modernization of building design has caused an increase in demand for quick indoor building environment simulations for the verification and optimization of the final design. Yet, the limiting factor for the accuracy of these simulations is the detail of the indoor geometry of the building in question. To provide an accurate level of detail, indoor environment modelers must rely on precise 3D geometries to depict a realistic performance of the air flow, envelope effects, or human interaction inside of the space (Privara et al. 2013). Ultimately, obtaining the correct building geometry provides for accurate model creation and precise simulation results. Recent methods for obtaining high-quality data rely on expensive sensors and human operators manually guiding the systems throughout a building. Not to mention, most research in this field focuses on using the obtained geometric data for robotic applications instead of modeling purposes (Francis et al. 2015). Current applications use various methods to capture the indoor and external geometry of a building, such systems include RGB depth cameras (RGB-D), light based sensors (LiDAR), and acoustic based mapping. All these systems rely on the same basic principles of Time-of-Flight (ToF) which itself has inherent disadvantages yet individually each of the systems have their own advantages. Finally, a ToF camera is a range imaging system that resolves distances based on the known speed of light or sound, and the measured time of flight of the signal sent to the object being pictured. This distance data is then processed as a point in 3D space (Pycinski et al. 2016).

As previously stated, ToF cameras suffer from several disadvantages such as interference due to external signals, distortions due to unwanted reflections, and background light (only applicable to light based systems) (He et al. 2017). Yet, the concept behind ToF cameras is still a prevalent method for obtaining distance measurement based images. For example, RGB-D cameras have recently come into play to build content rich 3D environments for robotic navigation, manipulation, telepresence, and semantic mapping. An RGB-D camera is a combination of a regular color camera with a depth camera which can provide RGB data along with pixel depth data. Nevertheless, regular depth cameras suffer from an inability to extract accurate depth data in indoor environments with dark or sparsely textured areas (Peter Henry 2010). Yet, by using, RGB-D cameras, researchers have been able to overcome these drawbacks and are able to obtain somewhat accurate representations of the space being analyzed. On the other hand, laser based, LiDAR, mapping has become a popular choice for mapping indoor environments. LiDAR works by illuminating a target surface with pulsed laser light and then waiting for the reflected signal. Similarly, LiDAR suffers from the regular ailments of ToF cameras, yet are very expensive when compared to an acoustic based ToF camera.

Likewise, acoustic based ToF cameras have become commonplace in the industry to map locations specifically for low cost robotics applications. Yet, the accuracy of this system still does not provide the level of 3D topographical detail which would be beneficial for indoor environment modeling applications but with the right processing, can potentially reach the details of an RGB-D camera for a fraction of the cost. Consequently, we propose a low-cost acoustic sensor array (less than $70) to automatically detect and map building geometry using echolocation as well as indoor room topography. Our focus of this research is to allow building modelers to obtain geometric information of existing buildings for use in indoor environment simulations. Nevertheless, for the experiments conducted in this study, the camera used consisted of an array of ten sonar sensors whose purpose was to calculate the distance of the device to the walls in a room. Finally, three cases were considered to provide points for comparison. In the first two cases, we study the feasibility of creating images out of the raw range data using image processing techniques. The techniques...
used in the algorithm were: interpolation, point spread function estimation and deconvolution, filtering using a Gabor filter, and edge improvements using Laplacian masking. Next, we study the mapping of a corner in a room, here the results are presented in the form of raw distance data overlaid onto the space. Finally, conclusions and further suggestions are made based on the results.

METHODS
The following sections introduce the ToF device, its downfalls, the theory behind ToF systems, and the processing algorithm used to process the data for the case studies.

ToF Device and Its Downfalls
The ToF device used in this study consisted of an array of sonar sensors controlled by a Micro-controller, the raw data was then sent to Raspberry Pi and processed using OpenCV. The parts used in the device were ten HC-SR04 ultrasonic range sensors (sonar sensors), one Arduino MEGA Microcontroller, and a pair of 180° rotation servos. Finally, these components were assembled in a fashion as seen in Figure 1. As one can see, the assembly consisted of pieces of wood to hold the system together, Table 1 shows the cost brake down for the unit.

![Echolocation device used for automatic geometry detection.](image)

Figure 1. Echolocation device used for automatic geometry detection.

<table>
<thead>
<tr>
<th>Part</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood / Wires / Misc.</td>
<td>$5</td>
</tr>
<tr>
<td>Arduino MEGA</td>
<td>$35</td>
</tr>
<tr>
<td>Servos</td>
<td>$10</td>
</tr>
<tr>
<td>HC-SR04 ultrasonic range sensors</td>
<td>$15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$65</strong></td>
</tr>
</tbody>
</table>

Table 1. Bill of Materials for Sensor System

From Figure 1, one can see the sensor array where each sensor has a separation of about 4-cm from each other. This was done to reduce the risk of cross talk between each ultrasonic sensor and avoid any false readings that may arise. Conversely, this creates the first problem in the overall output. By spacing each sensor, one is effectively creating missing data points. From our calculations, it was estimated that each 4-cm gap would create 152 missing pixels in the final image. Likewise, during the operation of the unit, the system automatically sweeps from left to right and up and down to “photograph” the area correctly. This again creates an artifact in the data, as the device sweeps through the azimuth plane, a gap of 1 cm is created which results in 36 missing pixels. The missing data is then corrected in the studies by an interpolation based on a Delaunay triangulation to match missing data.

Theory
Now, one must understand the basics behind sonar based distance measurements to understand the importance of post-processing the raw data. A distance measurement taken by the sonar sensors is done by sending a PING (a small burst of a high frequency signal) using its emitter and waiting for its echo on a receiver. Finally, once the echo arrives at the sensor, the time of flight, \( t \), (time it takes the echo to return) is multiplied by the speed of sound, \( v \), 340 m/s, then distance to the object is then given by Equation 1.

\[
d = \frac{vnAt}{2}
\]

One can assume that for most measurement calculations, the process described above works quite well. However, this is far from the truth. In practice, when taking measurements by using air-based sonar, certain factors come into play such as, interaction with the target (specular or surface reflections), variation of propagation speed due to temperature in the environment, and finally the opening angle of the transmitted beam. These issues affect the final distance measurement and in term affect the final range image.

The device used in this study, uses a PING that consists of a 40 KHz tone that is held for 40-uSec. Consequently, this type of beam is proficient for up close imaging yet it becomes very poor at further distances due to its far field pressure characteristics as we will see in the second case study. Thus, this is one of the major disadvantages of using sonar based ToF cameras. One way to fix this issue is by increasing the tone frequency of the PING. However, constructing a piezoelectric transmitter capable of such frequencies is quite expensive. Thus, in this study, by using image-processing techniques, this, and other distortions in the final image may be corrected without the need for expensive hardware-based solutions.

Processing Algorithm
As previously mentioned, the raw data of the ToF device has a variety of issues ranging from missing data values to distorted edges due to servo movement. However, the digital image-processing algorithm proposed below, we believe, can help alleviate these failures, Figure 2.

![Image Processing Flow Chart](image)

Figure 2. Image Processing Flow Chart

First, raw data comes from the Arduino Micro-controller. This information is composed of distances and locations...
calculated by the sonar. These distances have been translated from spherical coordinates to a Cartesian representation. One could consider this a type of "lens correction" for system. Then, the raw data from the sensor is interpolated by using the Delaunay Triangulation. This fills in the missing pixel values caused by the gaps between the sensors and the sweep angle. Nevertheless, the triangulation makes several distortions in the image which will then need to be corrected. Next, to decrease the blur in the image due to the triangulation and the motion of the sensor, the deconvolution of the point-spread function (PSF) was implemented. The point-spread function of the image can also be interpreted as the impulse response of the system. The PSF contains all the information about the artifacts introduced by the system. Therefore, in practice, one could describe the resulting output image of a system as follows, Equation 2.

\[ I_{\text{output}}(m,n) = H(m,n) + I_{\text{real}}(m,n) \] 

(2)

The equation above represents an unwanted convolution of the PSF with the original image. Here, \( H \) represents the PSF of the system and \( I_{\text{real}} \) describes the image without any distortions. Finally, \( I_{\text{output}} \) describes the distorted output image caused by the system. In practice, by translating Equation 1 into the Fourier domain one may manipulate it in such a way to obtain the original image, \( I_{\text{real}} \). The manipulation can be seen below, Equation 3. Notice that the \((m,n)\) have now become \((x,y)\) due to the Fourier transformation.

\[ \frac{I_{\text{output}}(x,y)}{H(x,y)} = f_{\text{real}}(x,y) \] 

(3)

Finally, in application, this process of distortion removal is not perfect. The biggest drawback comes from estimating the PSF for the system. Without an accurate PSF estimation, one might not fix the image or in the worst-case scenario, destroy it. However, if done correctly, deconvolution from an unwanted signal can reduce blurring in a shaky camera, and many other artifacts. In other words, the purpose of deconvolution is to reconstruct the signal, as it existed before the unwanted distortion convolution took place.

For this experiment, the PSF was estimated by emulating a technique used for microscopes (Zahreddine 2013). The technique calls for imaging an impulse like shape and then extracting the intensity values from the diameter of the imaged shape. Finally, the extracted intensity values from several trials are concatenated to form the estimated PSF. In the case of a microscope it calls for using beads with a 200-nm diameter. Yet, for this study, a circular block of wood measuring 10.16-cm in diameter was used instead. The circular shapes are of preference due to their likeness to an impulse in the Fourier domain.

Now, due to the ringing effects created by the deconvolution of the PSF, the next step to improve the image was to use a filter that removes such artifacts. For this, a Gabor Filter was used. Gabor filters exhibit several properties which conventional smoothing filters lack (Liu et al. 2014). The Gabor Filter is a composition of a Gaussian filter multiplied by a complex sinusoid. Mathematically, this can be expressed as, Equations 4-6.

\[ H(x,y) = s(x,y)g(x,y) \] 

(4)

\[ s(x,y) = e^{-\Delta x^2(u_x+v_y)} \] 

(5)

Equation 4, shows the composition for the filter, \( s \) represents the complex sinusoid, and \( g \) represents the Gaussian filter. Finally, the multiplication by the complex sinusoid is equivalent to translating the Gaussian function by \((u_x, v_y)\) in the frequency domain. Thus, this allows us to specifically target a direction that we wish to emphasize the edges in. This is useful for many types of applications such as, texture segmentation, edge detection, retina identification, and image representation. Due to this type of response, the Gabor filter can be described as a type of band pass filter. In other words, it will sharpen the edges in the direction that is desired while also smoothing the image. As previously discussed, to further improve the edge definition of the image a Laplacian Masking was used, which is a common technique for improving IR based images (Işık, Jane, and İlk 2011). The basics behind this mask involve taking the second derivative of the image. Once that has been done, at the places where the derivative is equal to 0, \( \Delta \) is where there is an edge in the image. Therefore, by adding this filtered version of the image back to the original, one effectively improves the edge distinction, this can be mathematically modeled as Equation 7.

\[ g(x,y) = f(x,y) + c[\Delta^2 f(x,y)] \] 

(7)

Here, \( g(x,y) \) represents the final sharpened image, \( f(x,y) \) is the original image and \( \Delta^2 f(x,y) \) represents its derivative. The factor of \( c \) is also there as a scaling factor for the derivatives contribution. However, for the mask used in this study a factor of -1 was used. Overall, the results for using the Laplacian mask were quite interesting as will be later seen in this paper.

**EXPERIMENTAL SETUP:**

**Case 1- Description:**

By using the ToF camera mentioned earlier in this paper, 3 range images were captured at different distances away from the sensor. Two of those images consisted of a block of wood, and the final image consisted of a wall in a room containing several objects. A regular image of the block of wood at the varying distances and the wall of a room can be seen below in Figure 3.

**Figure 3. Image of the Wooden Block Used at 0.5-m from sensor away (top-left), 0.15-m away from sensor (top-right) and the room wall, 1.83-m away from sensor (bottom).**
Implementation

The device was set up in front of the blocks of wood and started photographing the items by scanning up and down. In other words, the device started tilted up, it then would take a range measurement, and then move the device head down 5°. This process was done until the system had scanned the full object. During this time, the Arduino Micro-controller was connected via USB and the raw data was captured to a log file and transferred into the algorithm where the individual range data pixels were processed in real time. The processing algorithm was coded using C and implemented using the OpenCV library and ran on Raspberry Pi. The algorithm followed the flowchart found in Figure 2. An overview of the processing steps can be seen in Figure 4, here the steps show the processing for imagining a block of wood at 0.5-m away from the sensor.

![Figure 4. Process of the Image Processing Algorithm for the Wooden Block Used at 0.5-m away from the sensor.](image)

As one can see from Figure 4, the processing of the image is quite remarkable. Starting from the top left of the figure (section I), one can see that the range data is quite sparse and that interpolation must take place to compose a rational image. The next step in the chain is to create the interpolation between the data points. This is seen in the top middle of the figure (section II), likewise some of the artifacts mentioned before of the Delaunay triangulation used to estimate the interpolation causes a triangular distortion. Next, the PSF deconvolution is applied and the results can be seen in the top right of the figure (section III). Finally, the image is filtered using the Gabor filter and then adjusted for contrast, the results can be seen at the bottom left and bottom middle (section IV, V), respectively. The final results are displayed by color mapping the image, this can be seen at the bottom right of the figure (section VI). The results for all the three images in Figure 3 are further analyzed in the later sections of this study.

Case 2 - Description:

Following the first case, we focused the ability of the unit to mapping the dimensions of a space. This is the most beneficial aspect for the 3D modeling besides the internal topography of a space which can be captured using the techniques outlined and used in Case 1. The space in question for this test was chosen to be the corner of a room. Specifically, the system would be in-between a west wall, and a piece of furniture on the east and a north wall. The room in question can be seen in Figure 5. The walls are made of drywall with wooden studs found every 0.3-m. Finally, the piece of furniture is made from wood with a specular finish on the surface.

Figure 5. Room corner used for the room mapping experiment.

![Figure 5. Final Output Comparison of the Post-Processing System for a Block of Wood at 0.5-m Away from the sensors.](image)

**RESULTS**

This section explores the results for the two case studies. They have all been color mapped to provide more visual contrast. For visualization purposes, in Figures 5 and 6, the red portion in the output image represents objects closer to the sensors.

Case 1 - Results:

As mentioned before, for this case we imaged a block of wood at two different distances as well as a wall of a room. The block of wood was set at 0.5-m and 0.15-m away from the sensor. The comparison between the real image and the acoustically imaged piece for the 0.5-m and 0.15-m away blocks are found in Figures 5 and 6 respectively.

Figure 6. Final Output Comparison of the Post-Processing System for a Block of Wood at 0.15-m Away from the sensors.
By putting the images alongside the original pictures, one can better see how accurately the sonar sensors captured the objects. Also one must take note that the angle at which the original, regular, photographs were taken is as close to the viewing angle that the sensors had to the objects. On the other hand, Figure 6 exemplifies how the ToF camera has better lateral resolution at closer distances. We can see that the edges are more defined for the output image. As mentioned at the beginning of the paper this is due to the far field pressure of the 40kHz tone used to measure the ranges. In other word, the PING beam is narrower at closer distances thus providing better later resolutions.

Finally, an image of a room’s wall was taken. This proved to be the biggest challenge for the ToF camera. The comparison results can be seen below in Figure 7. Here, red symbolizes objects farther away from the sensors.

![Figure 7. Final Output Comparison of the Post-Processing System for Room Wall at 1.83-m Away from the sensors.](Image)

In this image, the back wall is 1.83-m away and the objects are around 0.91-m away from the sensor. Therefore, in the example of the room, one can really see the accuracy of the post processing system and the downfalls of using the sonar sensors. As a disclaimer, the final output of Figure 7 has been mosaicked due to the ToF camera’s physical limitations of only creating very directional, narrow images. Subsequently, the algorithm can define the edges of those objects that were closer in the image. For instance, if one looks closely to the left side of the output image for Figure 7, one will notice that the sensor is able to pick up the side edge of the desk, i.e. the blue vertical lines at the edge of the image represent the desk edge in the actual picture. Moving on to the middle of the image one can appreciate a large amount of red. This is due to the far away distance of the TV in the original image.

Finally, due to the reflective properties of the TV’s glass, it made the data acquisition by the sensor a lot more difficult thus the resulting shape, even after processing, is quite poor. As for the drawer in the original image, it suffered the most lateral definition distortion at the sonar output. One can see that in the processed sonar image, the drawer is almost nonexistent. The only trace in the sonar image of the drawer is at the middle-right of the image in a light green tint.

![Figure 8. Final Output Comparison of the Post-Processing System for Room corner obtained by the sensors.](Image)

**Case 2- Results:**

Finally, for Case 2 we explored the feasibility to use the sensor to map a space. In this case, the mapping is just a 2-D distance measurement which can then be used to extrapolate a 3-D space if combined with the topographical results of Case 1 and if measurements are taken at different heights. Nevertheless, the results for the mapping are overlaid onto the original room corner picture and can be seen in Figure 8. The red lines display the 2-D data collected by the sensor system.

From the acquired results in Figure 8 one can see that the North wall of the room corner is the best recognized by the system and can be accurately matched to the space. It is detected to be at about 40-cm away from the sensor which was placed on the south edge of the corner, at (0-cm, 0-cm). Nevertheless, if we look at the west wall of the corner or the east wall with the furniture one may notice that the red line diverges from the measurement and reflects up the wall or furniture. This is due to the angle at which the sensor sends the PING. Since the PING is quite directional the angle due to the rotation of the device creates a poor reflection when returned and this creates the artifact of a larger distance than what should be detected. Once again this displays the fact that sonar based ToF cameras are not very good due to interference or reflections.

**DISCUSSION**

From the results in the two cases studied in this paper, we see that the imaging accuracy for an acoustic ToF camera can be improved if the image is processed correctly and if the objects are close to the camera itself. Likewise, the 2-D results found during the second case of this study show that the distance can be accurately obtained by the sensors, however, due to the directionality of the PING unwanted artifacts due to reflections are created when the sensor is in mid rotation.

Also, If the 2-D distance data obtained in the second case is combined with the imagining date from the first case, we would be able to create a full system for mapping a 3D space. Unlike conventional methods being used in the industry like LIDAR and RGB-D the cost of the system is quite low considering the result obtained by both LIDAR and RGB-D are very like the results obtained here for the 2-D distance measurement. A LIDAR system can go for around $1,500 while a RGB-D system can cost around $150. Nevertheless, an RGB-D system can give more details than our acoustic imaging system and algorithm. Eventually, if our algorithm were to be tailored to be better during the interpolation process, we believe that the images obtained would be much more representative of the real space. Also, if we were to modify the existing system to reduce the distance between each of the sensor, then we could ideally reduce the number of points that must be interpolated and ultimately increase pixel density and overall image definition.
Physical Property Interaction

During the study, an idea to obtain the imaged objects physical properties based on the reflected signal strength was considered. However, this was not possible due to the sensors inability to output raw intensity values. Nevertheless, an image was taken of a piece of glass to show how its high-reflectance affected the final edge definition.

Figure 9. Final Output Comparison of the Post-Processing System for a Glass Sheet at 0.3-m Away from the sensors.

From Figure 9 one can see that it looks quite like the previous examples done with the blocks of wood. Nonetheless, one should see how there is a wider green halo and more changes in the background around the solid blue block. This is due to the scattered reflection of the PING from the glass. Instead of a smooth background, the sensors picked up more “garbage” data from the echoes, even if the sensor was not pointing directly at the glass. Lastly, one can see that the sensors did not detect the base holding the glass. Instead of a sm

post-processing techniques used, did help improve the overall result of the final image. The best example of this was Figure 6. In that image, it is easy to appreciate well-defined edges undisrupted by the poor lateral resolution of the sensors. Likewise, in Figure 4, one can see the full power of concatenating different techniques into producing a final image from the sonar range data. All in all, even though the poor lateral resolution of the sensors and the device distortions can’t be fully rectified by the post-processing system, they can be reduced a considerable amount, enough to produce usable data. Likewise, we believe that even though the results for either of the two cases are not perfect in representing a space. The combination of the two features in future studies could provide a path for creating a portable, affordable, solution for quick building mapping to be used in interior building modeling. Specifically, we find our system to be valuable to architects or engineers who require the mapped data to plan retrofits or renovations without investing time and money in more expensive solutions. Some other examples which require 3D indoor models include indoor airflow simulation using Computational Fluid Dynamics, disaster management simulation studies, and interior design tools.

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation under Award No. IIS-163333. Matlab_R2014b was used to provide the mosaics for some figures.

REFERENCE


Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, Dieter Fox. 2010. 'RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments'.


Zahreddine, Ramzi N ; Cormack, Robert H ; Cogswell, Carol J. 2013. 'Noise removal in extended depth of field microscope images through nonlinear signal processing', Applied Optics.