Fit My Ears: 3D Reconstruction of Ear Geometry from Photographs for Personalized Audio and Hearing Protection

Zamir Ben-Hur
zamirbh@stanford.edu

Paul Calamia
pcalamia@standford.edu

Abstract

Knowledge of the 3D geometry of human ears is valuable for a number of research fields including biometrics and personalized spatial audio due to each ear's unique shape. In addition, the improved fit of personalized earbuds and earplugs, form-fit to the user's ear shape, can provide increased comfort and benefits in audio delivery and hearing protection, respectively, through improved sealing of the ear canal. Because 3D ear scans are impractical to use at scale, but high-resolution ear photographs and videos are trivial to create with cell-phone cameras, this paper explores the use of several photogrammetry pipelines for 3D ear reconstruction. Up through our final presentation, our results, using both photographs of real ears and images of 3D ear models, suggested that modest ear reconstructions could be made with approximately 100 images using open-source structure-from-motion (SFM) tools, but we had not achieved results that were sufficient for creating personalized in-ear devices. Since then we began working with a new SFM tool and now have a limited set of results with significantly improved reconstruction performance using as few as 23 images, which is comparable to other published ear photogrammetry results.

1. Introduction

Knowledge of the 3D geometry of human ears is valuable for a number of research fields including biometrics [31] and personalized spatial audio [32] due to each ear's unique shape. For example, for creating accurate spatial-audio playback, which is important for immersive virtual- and augmented-reality applications [10], personalized head-related transfer functions (HRTFs) are required. Such HRTFs can be generated by computer simulations using a 3D model of the person's ear [16]. In addition, the improved fit of custom earbuds and earplugs, form-fit to the user's ear shape, can provide increased comfort and benefits in audio delivery and hearing protection, respectively, through improved sealing of the ear canal [28, 30].

Traditionally, 3D ear geometry is acquired using specialized and expensive equipment, such as ear molds and 3D scanners (more details in Sec. 2), which is impractical to use at scale. With the proliferation of high resolution cameras on mobile devices, such as cell-phones, 3D reconstruction of the ear from such cameras' photos and videos is a potentially more viable solution.

In an effort to better understand the potential for personalized in-ear devices at scale, in this work we intend to evaluate various photogrammetric pipelines that take as input a collection of photographs of a listener's ear and output a 3D reconstruction. Such a reconstruction can be used to create personalized HRTFs or customized tips for earbuds and in-ear hearing protectors, by, for example, extracting the ear dimensions as seen in Fig. 1. As input data we have been using photographs and video frames, as well as images of 3D ear models (see details in Sec. 3). We have evaluated several pipelines for 3D ear reconstruction from 2D images using existing structure-from-motion (SFM) and multi-view-stereo (MVS) pipelines, with the aim of comparing their 3D reconstruction performance and finding the optimal number of required input images.

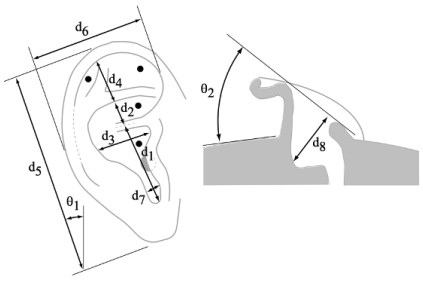


Figure 1. Ear dimensions that can be used for personalized HRTF prediction or customized ear-tips. Adopted from [9].

2. Background and Related Work

Historically, custom-fit in-ear devices have been designed using molds of the outer ear (specifically the concha) and the ear canal, but recently various 3D scanning devices have become popular. For example, The Lantos ear-

scanning system [6] reconstructs ear-canal geometry from a series of photographs taken by a movable camera inserted into an inflatable in-ear membrane, and the Otoscan system [8] employs a "ring laser scanner" for similar reconstructions.

Research on personalized spatial audio also has explored multiple technologies for generating 3D ear models which can be used in conjunction with boundary-element, finite-element, or finite-difference time-domain methods to generate computational simulations of HRTFs. Recent work has included the use of structured-light 3D scanning [13, 32], MRI scanning [11], and SFM pipelines which process photographs or video frames from digital SLR cameras [20] and cell-phone cameras [18], along with comparative studies of different scanning techniques [16, 24]. Several audio-technology companies currently offer cloud-based HRTF simulations based on images and/or videos from cell-phone cameras [4, 5].

3. Approach

Our evaluation approach leveraged multiple implementations of photogrammetric 3D reconstruction algorithms available for use without purchase, specifically OpenMVG [21] with OpenMVS [14], COLMAP [26, 27], Meshroom [7] and Agisoft Metashape [1][1].

OpenMVG is an open-source SFM implementation that offers both sequential and global SFM pipelines which can be coupled with dense reconstruction from OpenMVS to produce point clouds and meshes from sets of images. At the time of writing we have been unable to get any useful results from OpenMVG after multiple attempts so it will not be discussed further. COLMAP is also an open-source SFM/MVS implementation with a basic pipeline for feature detection/extraction, feature matching across images, and bundle-adjustment-based structure and motion reconstruction. Similarly, Meshroom is a free (but not open-source) incremental SFM implementation. Further details about the included algorithms and options can be found in the respective documentation and web pages. Agisoft Metashape is a closed implementation of SFM which is based on a mix of published and proprietary algorithms. All software was installed and run on a Falcon Northwest Talon desktop computer with a 3.5 GHz Intel Core™ i7-7800X CPU, 32GB of RAM, an NVIDIA GeForce RTX 2080 GPU, and Windows 10. All of the reconstruction programs were run with their default settings.

As described in more detail in Section 4, for input data we have used photographs from the AMI Ear Database for biometric ear identification [2] (Section 4.1.1), photographs taken by the authors (Section 4.1.2), and images of 3D ear models from the Children's HRTFs and Anthropometric Scans for Auditory Research (CHASAR) database [3, 13] (Section 4.2). Ultimately our system must work with photographs, but a key disadvantage to using photographs as input at this stage is that there is no 3D ground truth to utilize for performance evaluation. There is also no simple way to recover scale from non-annotated photos. While images of digital ear models (specifically, 3D meshes) may suffer from unrealistic lighting and/or skin textures, they offer the benefits of unlimited numbers of views, ground-truth geometry for reconstruction validation, and easy annotations for scale references and to aid in finding correspondences.

Because our ability to obtain high-quality reconstructions came only in the last few days, this report contains only qualitative analyses of our results. As described below in Section 5 various metrics are available for assessing the quality of 3D photogrammetric reconstructions quantitatively including model accuracy and completeness [22,24, 25,29], and the Hausdorff [19,23] or chamfer [12] distances between ear point clouds.

4. Experiments

As is mentioned above and is discussed in the following subsections, we processed images from a publicly available dataset of ear photographs (Sec. 4.1.1), ear photographs that we took (Sec. 4.1.2), and images of 3D models (triangle meshes) of ears (Sec. 4.2). For the image datasets our experiments mainly focused on the 3D reconstruction performance as a function of software/implementation and as a function of the number of input images. For the 3D-model data we explored the same parameters as well as the effects of varying the lighting and skin texture, and the use of colored markers to aid in the feature detection/correspondence process.

4.1. Photographs

4.1.1 The AMI Ear Database

Our first attempt at ear reconstruction utilized photographs from the AMI Ear Database [2]. This database includes six images of the right ear from 106 subjects, all taken from a distance of approximately 2 meters. Five of the images were taken using a focal length of 135mm with the subject asked to move their head in a series of prescribed motions (labeled 'front', 'up', 'down', 'left', and 'right'). The sixth image is a zoomed version of the 'front' image taken with a focal length of 200mm. An example of the six images for Subject 000, along with 3D reconstructions (point clouds) created with the 'Automatic Reconstruction' tool in COLMAP and with Agisoft is shown in Figure 2. Meshroom was not able to process this collection of photos but the limited diagnostic messages made it difficult to troubleshoot.

As can be seen in Figure 2 (g) and (h), the reconstruc-

[1] Agisoft Metashape is proprietary software which typically requires the purchase of a license, but we obtained a free, temporary license from https://agisoft.freshdesk.com/support/solutions/articles/31000164657.

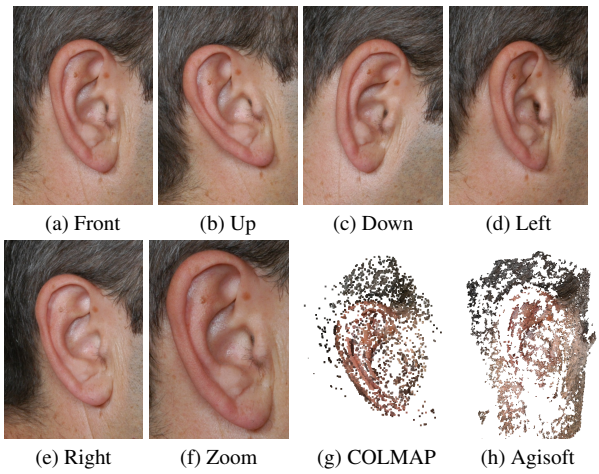


Figure 2. Ear images from Subject 000 in the AMI Ear Database ((a) - (f), captions indicate the subject's look direction) and 3D reconstructions from COLMAP (g) and Agisoft (h).

tions are recognizable as an ear, but are far too incomplete to be usable for our application. Subsequent tests using the images from 7 randomly chosen subjects from the AMI database yielded only the poor reconstructions for Subject 032 shown in Figure 3. Otherwise no further reconstruction results were produced. Whether this is due to different lighting conditions or some other factor is unclear, but given the poor results from Subjects 000 and 032 we have concluded that the small number of photos and the limited variety of poses in this dataset are insufficient for accurate 3D reconstruction.

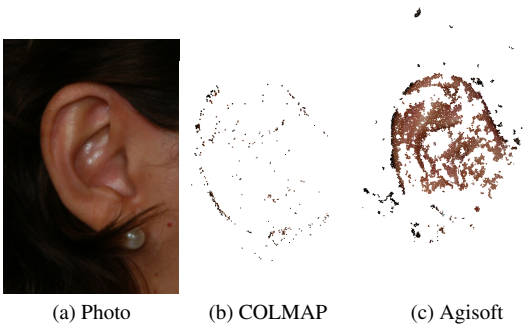


Figure 3. 3D Reconstruction results for AMI Subject 032 using COLMAP and Agisoft.

4.1.2 Our Photographs

To test our hypothesis that the limited number of viewpoints in the AMI dataset caused the reconstruction failure, we needed a dataset with many more photographs. Nothing appropriate was available to download, so we collected photographs of a right ear from one subject, taken at a variety of

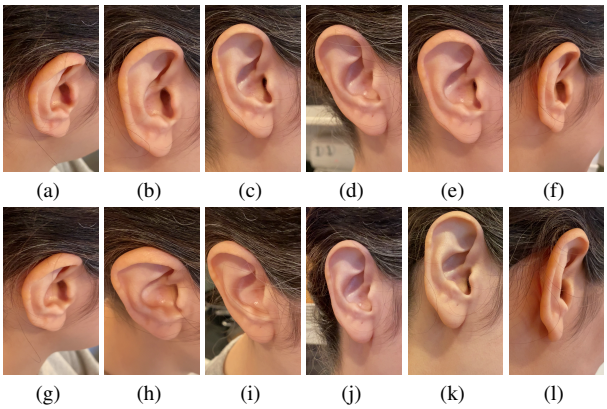


Figure 4. Ear images of subject 001 taken from video frames captured by the authors, using an IPhone 11 Pro.

azimuth and elevation angles. A rapid way to collect such data is by capturing a video and extracting photographs as frames of the video. We captured a 1 minute, 30fps video using an IPhone 11 Pro and extracted 91 photographs from various camera orientations. A subset of 12 of these photos is shown in Figure 4.

Figure 5 shows the reconstruction results using the Agisoft implementation with all 91 photographs. The figure contains both the point-cloud and the output 3D mesh compared to an input photograph. It can be seen that, qualitatively, the reconstruction is quite good. The general shape of the pinna is well observed as well as its fine structure, and most of the dimensions presented in Fig. 1 can be easily extracted (but additional information is required to obtain the proper scale).

Figure 6 shows a comparison of the reconstructed point clouds using the three software implementations (COLMAP, Meshroom and Agisoft) and all 91 photographs. It can be seen that Meshroom and Agisoft were able to reconstruct the ear much better than COLMAP, which yielded significant gaps and noticeable depth discontinuities. Agisoft achieved a higher point density than Meshroom, but further analysis is necessary to determine how much of an effect that might have on using the reconstruction for an in-ear device fitting.

To test how many photographs are required for reasonable reconstruction, we took subsets of photos from the 91-photo set, with 12, 23, and 46 photographs. Figure 7 shows the reconstructed point-clouds using Agisoft with the various subsets. As can be seen in the figure, with as few as 23 photos the reconstruction still looks reasonable. Although some missing points are visible, the geometry of the ear is clearly observed. With 12 photos the result is severely degraded, with significant gaps in the ear surface.

(a) Photograph (b) Point cloud (c) Mesh

Figure 5. Reconstruction with Agisoft using 91 photos.

(a) COLMAP (b) Meshroom (c) Agisoft

Figure 6. Point-cloud reconstructions for our subject 001 from 3 different software implementations with 91 input photos.

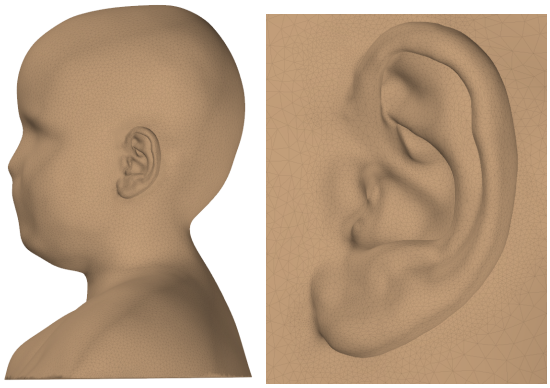
(a) 12 (b) 23 (c) 46 (d) 91

Figure 7. Point-cloud reconstructions with Agisoft using increasing subsets of photos from our 91-photo set for subject 001.

4.2. Rendered Images from a 3D Model

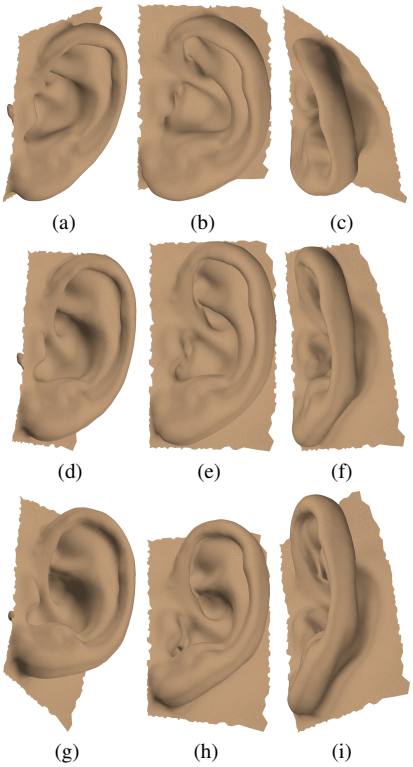
As mentioned in Section 3, it was clear that our 3D reconstructions from photographs would be difficult to evaluate quantitatively due to a lack of ground-truth geometry. This challenge led to the exploration of the use of images of 3D ear models as the input to the photogrammetry pipelines.

Our models were obtained from the CHASAR database, a collection of head and upper-torso meshes created for numerical simulation of HRTFs [3, 13]. An example of a full model and a cropped left ear are shown in Figure 8. A typical model in the dataset has approximately 115k triangles and 57k vertices, of which approximately 28k triangles and 14k vertices make up each ear and a small surrounding area. All images were created in Matlab by loading in the associated geometry, cropping to a small region around an ear, and rendering the image at a series of view directions with skin texture, colored markers (to aid finding correspondences), and lighting as described below. An example set of 9 views, in a 3x3 grid spanning 90° in azimuth and 60° in elevation (relative to a view aligned with the interaural axis) is shown in Figure 9.



(a) Full Model. (b) Cropped left ear.

Figure 8. CHASAR Model 01



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 9. 9 views of the left ear from the CHASAR model 01.

4.2.1 Effect of Image Count and Distribution

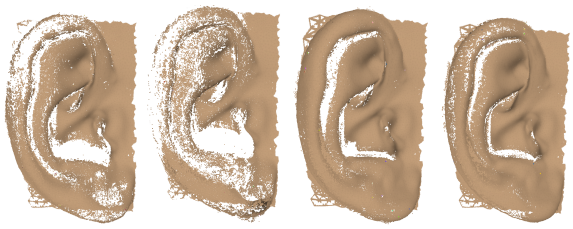
Given our results with photographs, as described in Section 4.1 and illustrated in Figure 7, which suggest that somewhere between 20 and 100 images is sufficient for a good reconstruction, we experiments with different numbers and arrangements of input images. For most cases we found that, for a fixed number of images, a regular grid produced better results than a random distribution. In the interest of conserving space, no results with random spacing are shown. We also found that image sets with cardinality less than approximately 50 led to clearly unusable results. Examples of some of our results are shown in Figures 10 and 11. Evenly spaced views, spanning 120° in azimuth and 90° in elevation relative to the interaural axis, in grids of 8x8, 9x9, 10x10, and 12x12, were used to create input image sets for COLMAP and Agisoft.

For the COLMAP results in Figure 10, performance is not monotonic with increasing image-set size; in particular the 9x9 grid yielded the worst reconstruction. For all reconstructions there are portions of the concha, the bowl-like portion of the ear that lead to the ear canal, which are missing. This would be a critical problem for our intended application as this is precisely the location where in-ear devices must fit well for comfort and proper sealing. The fossa and the underside of the helix (the surfaces under the outermost, tightly curved, portion of ear) are also poorly reconstructed, which would be problematic if the models were meant for HRTF simulations but acceptable for in-ear device fitting given the lack of device contact in those areas. Also note that all of the COLMAP results are inverted in the z-axis (depth), such that images of a left ear resulted in models that are oriented like a right ear

For the Agisoft results in Figure 11), the performance for the 8x8 and 9x9 grids *appears* to be improved over that from COLMAP, but rotated views of the reconstructions (not shown) show the points on the concha extruded outward from the ear rather than inward. This is also true to a lesser extent for the 10x10 grid. The point clouds for the 10x10 and 12x12 grids are surprisingly sparse, but as shown in Fig 11 (e) the meshed reconstruction for the 12x12 grid is quite good despite some surface roughness. Note that the 12x12 grid of images was the only case for which the orientation of the reconstruction is correct.

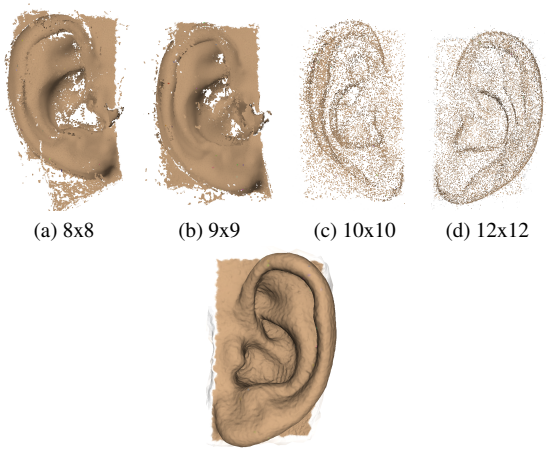
4.2.2 Effect of Lighting

We tried two different lighting schemes for our 3D models. The first (denoted 'Lighting 1' in Fig. 12) included the default Matlab light located at infinity along a [1 1 1] vector from the center of the image (with the positive x axis pointing forward toward the face, positive y out along the interaural axis, and positive z pointing up) augmented with a second light at infinity along the vector [-1 1 -1].



(a) 8x8 (b) 9x9 (c) 10x10 (d) 12x12

Figure 10. Point-cloud reconstructions of CHASAR Model 01 using COLMAP with different image grids spanning $[-60^\circ, 60^\circ]$ in azimuth and $[-45^\circ, 45^\circ]$ in elevation relative to the interaural axis.



(a) 8x8 (b) 9x9 (c) 10x10 (d) 12x12

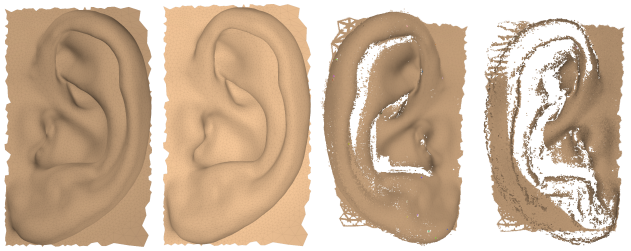
(e) 12x12 mesh

Figure 11. Point-cloud reconstructions, (a) - (d), and a mesh reconstruction, (e), of CHASAR Model 01 using Agisoft with different image grids spanning $[-60^\circ, 60^\circ]$ in azimuth and $[-45^\circ, 45^\circ]$ in elevation relative to the interaural axis.

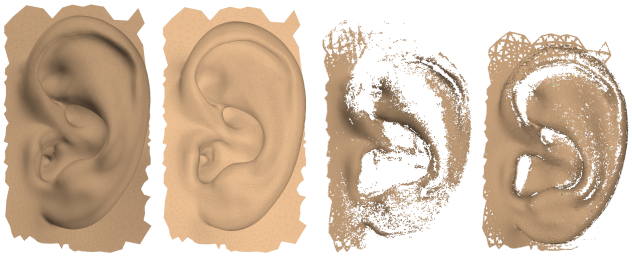
In Matlab the lights are fixed with respect to the coordinate system of the model, so rotating the model to change the view is equivalent to moving a camera around a fixed object with fixed lighting. Our second lighting approach (denoted 'Lighting 2' in Fig. 12) involved 5 lights whose orientations were fixed with respect to the camera (along the camera axis, $\pm 65^\circ$ in azimuth, and $\pm 65^\circ$ in elevation. The lights moving with the camera tended to generate more dynamic shadows. Changing the lighting did not have a consistent effect on reconstruction performance, as can be seen in Figure 12 with examples from CHASAR Models 01 and 22. COLMAP performed better with the model-fixed lighting on Model 1 but better with the camera-fixed lighting on Model 22.

4.2.3 Effect of Texture and Markers

The CHASAR models do not include any color or texture information. The baseline image-rendering parameters we used included Gouraud face lighting (which computes light



(a) Ground Truth, Lighting 1 (b) Ground Truth, Lighting 2 (c) Point Cloud, Lighting 1 (d) Point Cloud, Lighting 2



(e) Ground Truth, Lighting 1 (f) Ground Truth, Lighting 2 (g) Point Cloud, Lighting 1 (h) Point Cloud, Lighting 2

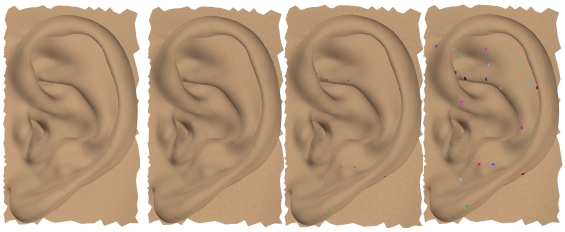
Figure 12. Point-cloud reconstructions of CHASAR Model 01 (top row) and Model 22 (bottom row) using COLMAP with evenly spaced 10x10 image grids and 2 different lighting schemes.

values at each model vertex and interpolates these across the faces, recommended within Matlab for curved surfaces), the dull material setting to remove specular highlights, and a Caucasian skin color with RBG values [228 185 142]. To add some texture to the ear surfaces we experimented with rendering the model edges, using with the same color as the faces. The visible edges significantly improved the reconstructions results (comparison not shown) so all reconstructions in this report were created with this option turned on. A comparison of CHASAR model 15 with invisible and visible edges can be seen in Figure 13 (a) and (b).

As mentioned in Section 3, one advantage of working with rendered images of 3D models is that we can annotate them arbitrarily to facilitate improved correspondence finding. We explored this by rendering images with and without randomly colored markers at randomly chosen vertices. Examples are shown in Figure 13 (c) and (d) with markers of 2 different sizes. The presence of markers improved the reconstruction performance but the size had no pronounced effect (not shown) so all reconstructions in this report were created using 32 small markers distributed over each ear model.

4.2.4 Effect of Software

Unlike the reconstructions from photographs of real ears where Agisoft and Meshroom provided superior results over COLMAP, we generally found that COLMAP provided the best reconstructions from images of 3D models



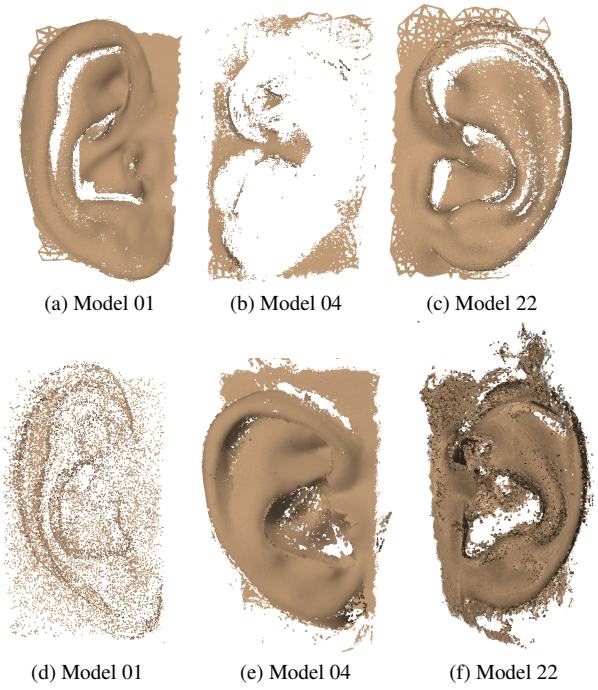
(a) No Edges, No Markers (b) With Edges, No Markers (c) With Edges, Small Markers (d) With Edges, Large Markers

Figure 13. CHASAR Model 15 with different skin textures and markers for correspondence finding.

(but still of insufficient quality for our project). Figure 14 depicts COLMAP and Agisoft reconstructions for 3 different models from the CHASAR dataset. In general, Meshroom was not able to process the 3D model images due to an inability to detect sufficient image features. For Model 01, the Agisoft point cloud is sparse but provides good coverage of the ear geometry while the COLMAP point cloud is significantly denser but has significant gaps. For Model 04 the COLMAP reconstruction failed and produced mainly points on the flat surface around the ear. The Agisoft reconstruction looks significantly better from this view, but has an incorrectly extruded concha as was seen with the photos in Figure 11 and described in Section 4.2.1. The results for Model 22 are similar in density and coverage but both have problematic gaps.

5. Conclusions and Future Work

In this project we explored the use of multiple SFM/MVS implementations to create 3D models of ears from 2D images to support the design of personalized in-ear devices such as earbuds and earplugs. Using photographs of real ears, we found that approximately 20-40 images are required for usable results, and that frames from a short 30fps video from an iPhone are sufficient. Agisoft Metashape and Meshroom provided high-quality reconstructions, while COLMAP was unable to match their performance. Our results are comparable to those in recently published work: Nightingale *et al.* [22] and Ross *et al.* [25] suggest that 30 photos from an iPhone are sufficient for a reconstruction with 71 $\pm$ 14% surface completion (relative to a model derived from a laser scan); Furferi *et al.* [17] suggest that 42 images (21 images spanning 200° in azimuth at 2 unspecified heights) with an iPhone are sufficient for reconstruction with error on the order of 0.5mm with respect to a depth-camera reconstruction, and Di Giusto *et al.* [15] found that "around 100" video frames were sufficient to reconstruct the ear of a dummy head with a maximum deviation of 2.5mm from the ground truth geometry.



(a) Model 01 (b) Model 04 (c) Model 22

(d) Model 01 (e) Model 04 (f) Model 22

Figure 14. Three left-ear models from the CHASAR database. Top row: COLMAP reconstructions. Bottom row: Agisoft reconstructions.

Because our ear photos had no ground-truth geometry, we also experimented with reconstructing ear geometry from rendered images of 3D models (triangle meshes). For these, we were unable to match the performance that we saw on the reconstructions from photos, with particularly problematic failures to reconstruct the geometry of the concha where in-ear devices are placed. Roughly 100 images were required for the best reconstruction results, which were obtained with COLMAP, and skin texture (in the form of visible mesh edges) and colored markers at randomly chosen vertex locations were also required. The reasons for this relatively poor performance is unclear, but it is likely that unrealistic lighting and texture were to blame so future work should be directed toward applying more realistic rendering techniques.

The main shortcoming of this work is that we were unable to generate results that were amenable to quantitative performance analysis. Related work has employed model accuracy and completeness [22, 24, 25] to evaluate ear reconstructions, where, given ground-truth geometry G and a reconstruction R, accuracy quantifies how close R is to G and completeness quantifies how much of G is modeled by R [29]. The Hausdorff distance also has been used for this purpose [19, 23], and the chamfer distance is another alternative for finding the distance between ear point clouds [12]. All of these should be explored in the context of personalizing in-ear devices to find which among them is most appropriate and what error threshold is acceptable for a comfortable and performance-enhancing fit.

References

[1] Agisoft Metashape Professional. https://www.agisoft.com/. [Online: accessed 15-March-2022].

[2] AMI Ear Database. https://ctim.ulpgc.es/research_works/ami_ear_database/. [Online: accessed 30-January-2022].

[3] Children's HRTFs and Anthropometric Scans for Auditory Research (CHASAR). http://publications.rwth-aachen.de/record/821483. [Online: accessed 30-January-2022].

[4] dolby atmos personalized rendering.

[5] Genelec Aural ID. https://www.genelec.com/aural-id. [Online: accessed 16-March-2022].

[6] Lantos Ear Scanning System. https://www.lantostechnologies.com/. [Online: accessed 16-March-2022].

[7] Meshroom. https://alicevision.org/meshroom. [Online: accessed 26-February-2022].

[8] Otoscan - 3D Ear Scanning Solution. https://hearing-balance.natus.com/en-us/products-services/otoscan. [Online: accessed 16-March-2022].

[9] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pages 99–102. IEEE, 2001.

[10] D. R. Begault. 3–D sound for virtual reality and multimedia. *NASA, Ames Research Center, Moffett Field, California*, pages 132–136, 2000.

[11] R. Bomhardt, M. de la Fuente Klein, and J. Fels. A high-resolution head-related transfer function and three-dimensional ear model database. In *Proceedings of Meetings on Acoustics, Acoustical Society of America*, volume 29, page 050002, 2016. Data available at https://www.akustik.rwth-aachen.de/go/id/lsly/lidx/1.

[12] G. Borgefors. Distance transformations in arbitrary dimensions. *Computer Vision, Graphics, and Image Processing*, 27(3):321–345, 1984.

[13] H. S. Braren and J. Fels. Towards child-appropriate virtual acoustic environments: A database of high-resolution HRTF measurements and 3D-scans of children. *International Journal of Environmental Research and Public Health*, 19(1):324, 2022.

[14] D. Cernea. OpenMVS: Multi-view stereo reconstruction library. 2020.

[15] F. Di Giusto, S. van Ophem, E. Deckers, and W. Desmet. Evaluation of the accuracy of photogrammetry for head-related transfer functions acquisition using numerical methods. *Fortschritte der Akustik-DAGA 2021*, pages 1614–1617, 2021.

[16] M. Dinakaran, F. Brinkmann, S. Harder, R. Pelzer, P. Grosche, R. R. Paulsen, and S. Weinzierl. Perceptually motivated analysis of numerically simulated head-related

transfer functions generated by various 3D surface scanning systems. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 551–555. IEEE, 2018.

[17] R. Furferi, E. Mussi, M. Servi, F. Uccheddu, Y. Volpe, and F. Facchini. 3D acquisition of the ear anatomy: A low-cost set up suitable for the clinical practice. In *Mediterranean Conference on Medical and Biological Engineering and Computing*, pages 669–678, 2019.

[18] T. Huttunen, A. Vanne, S. Harder, R. R. Paulsen, S. King, L. Perry-Smith, and L. Kärkkäinen. Rapid generation of personalized HRTFs. In *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, 2014.

[19] A. Mäkivirta, M. Malinen, J. Johansson, V. Saari, A. Karjalainen, and P. Vosough. Accuracy of photogrammetric extraction of the head and torso shape for personal acoustic HRTF modeling. In *Audio Engineering Society Convention 148*, 2020.

[20] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 53–61, 2014.

[21] P. Moulon, P. Monasse, R. Perrot, and R. Marlet. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74, 2016.

[22] R. C. Nightingale, M. T. Ross, R. L. Cruz, M. C. Allenby, S. K. Powell, and M. A. Woodruff. Frugal 3D scanning using smartphones provides an accessible framework for capturing the external ear. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 74(11):3066–3072, 2021.

[23] K. Pollack, P. Majdak, and H. Furtado. Evaluation of pinna point cloud alignment by means of non-rigid registration algorithms. In *Audio Engineering Society Convention 150*, 2021.

[24] A. Reichinger, P. Majdak, R. Sablatnig, and S. Maierhofer. Evaluation of methods for optical 3-D scanning of human pinnas. In *2013 International Conference on 3D Vision-3DV 2013*, pages 390–397, 2013.

[25] M. T. Ross, R. Cruz, T. L. Brooks-Richards, L. M. Hafner, S. K. Powell, and M. A. Woodruff. Comparison of three-dimensional surface scanning techniques for capturing the external ear. *Virtual and Physical Prototyping*, 13(4):255–265, 2018.

[26] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[27] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[28] B. Schulein and B. Fligor. In situ subjective and objective acoustic seal performance tests for insert earphones. In *Audio Engineering Society Convention 141*, 2016.

[29] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528, 2006.

[30] J. Stefanson and W. A. Ahroon. Evaluation of custom hearing protection fabricated from digital ear scanning and traditional methods. Technical report, US Army Aeromedical Research Lab, Fort Rucker AL, 2019.

[31] Z. Wang, J. Yang, and Y. Zhu. Review of ear biometrics. *Archives of Computational Methods in Engineering*, 28(1):149–180, 2021.

[32] Y. Zhou, H. Jiang, and V. K. Ithapu. On the predictability of HRTFs from ear shapes using deep networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 441–445, 2021.