https://doi.org/10.2352/EI.2023.35.17.3DIA-108

Assistive Mobile Application for Real-time 3D Spatial Audio Soundscapes Toward Improving Safe and Independent Navigation

Broderick S. Schwartz and Tyler Bell*

Department of Electrical and Computer Engineering, University of Iowa; Iowa City, Iowa 52242, USA
* ***tyler-bell@uiowa.edu***

Abstract

Assistive technologies are used in a variety of contexts to improve the quality of life for individuals that may have one or more vision impairments. This paper describes a novel assistive technology platform that utilizes real-time 3D spatial audio to aid its users in safe and efficient navigation. This platform leverages modern 3D scanning technology on a mobile device to digitally construct a live 3D map of a user's surroundings as they move about their space. Within the digital 3D scan of the world, spatialized, virtual audio sources (i.e., speakers) provide the navigator with a real-time 3D stereo audio "soundscape." As the user moves about the world, the digital 3D map, and its resultant soundscape, are continuously updated and played back through headphones connected to the navigator's device. This paper details (1) the underlying technical components and how they were integrated to produce the mobile application that generates a dynamic soundscape on a consumer mobile device and (2) a methodology for analyzing the usage of the application. It is the aim of this application to assist individuals with vision impairments to navigate and understand spaces safely, efficiently, and independently.

Introduction

The number of people that are blind or vision impaired is increasing globally, particularly as demographics shift and populations age [1]. For these individuals that are blind, severely vision impaired, or newly vision impaired (referred to collectively as people who are visually impaired, or PVI), safely learning and efficiently navigating a new environment may prove to be challenging. These challenges can be reduced by assistance from another being (e.g., sighted guide, guide dog, or remote assistant) or an assistive device (e.g., cane, talking GPS, sonar guide). Relying on another individual, however, may introduce a dependency which limits practical independence. Additionally, modern assistive technologies often give feedback to their users with low resolution, potentially making them unsuitable for a wide variety of situations. Recently, technologies that combine mobile imaging with virtual reality techniques have been successful in offering magnified views to low vision users [2]. However, such implementations may not improve visual motor function or mobility [2], may be relatively expensive, and may fatigue the user if worn for too long; it is also one more thing that PVI must carry with them.

Prior work has been conducted that encodes one's spatial surroundings into sound. For example, in [3], audio signals were used to alert and potentially guide the user. Another work in this area incorporated two cameras into a stereovision, head-mounted system worn by the user [4]. The two cameras worked together to reconstruct the 3D scene in front of the user. A "sonification algorithm" was then used to encode the 3D scene into an audio signal that could be played back to the user. While its users were able to employ the developed prototype for spatial orientation and obstacle avoidance [4], the head-worn device was somewhat bulky and required a 10-m cable tether to a PC or a special laptop backpack to be worn, thus limiting its practical viability.

In 2017, researchers incorporated a depth imaging camera (for indoor and low light settings) and a stereovision system (for outdoor settings) directly into specially designed headgear [5]. These imaging systems were then used to perform 3D reconstruction, 3D ground detection and segmentation, and ultimately detection of objects within the environment. While it provided an advantage by extending the scope in which such devices could be used, the user was still required to wear a tethered headset. This work [5] also notes reasons that consumer grade assistive systems have not seen wide adoption by PVI, such as form factor and the lack of efficient training programs.

Recent research [6, 7, 8, 9] has used computer-generated virtual environments to validate participants' ability to use spatial audio as a kind of "echolocation" to successfully learn and navigate the digital environments. As it relates to the approach presented in this paper, if an application can capture and reconstruct a user's physical surroundings in real-time, spatial audio signals (digitally placed within the reconstruction) may be able to provide users with the ability to safely learn and navigate their real-world environments. Irrespective of the many types of apparatuses and form factors available, there are some PVI who may not want to use an obviously visible assistive technology, which may mark them as "other" and create social barriers [10]. Thus, one additional challenge in developing a technology that utilizes spatial audio for navigation purposes is doing so within a form factor that is both lightweight and indistinct; potentially using devices that are already being carried daily.

The objective of this paper is to describe a mobile assistive technology that creates a 3D spatial audio "soundscape" to enhance the ability of PVI to understand and navigate spaces safely, efficiently, and independently. To increase the accessibility of this assistive technology, this work also aims to be compatible with devices that many PVI already own, such as an iPhone and stereo headphones. In general, the potential of this technology can have

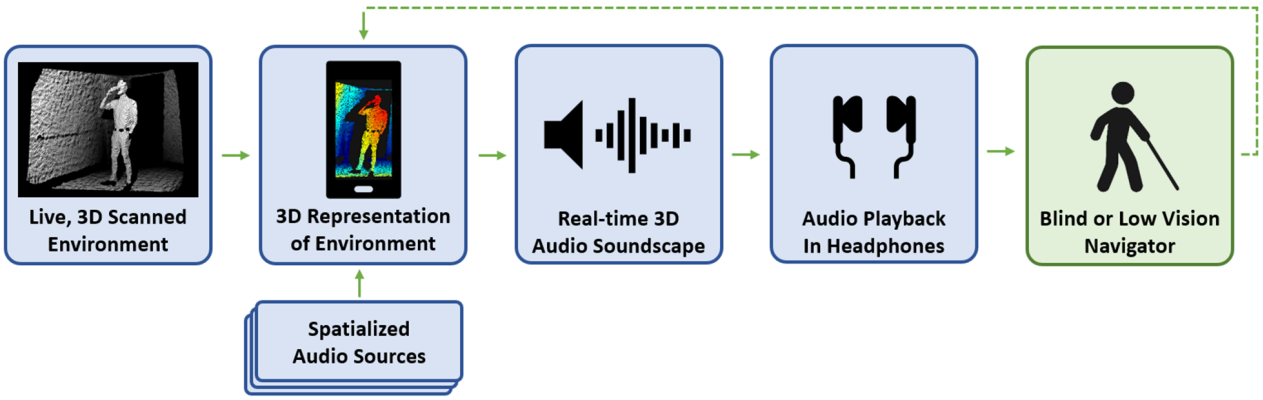

Figure 1. Block diagram of the proposed mobile application. First, the initial environment scan is made, creating a 3D map of the scene. Second, the 3D map is processed using the underlying game engine and spatialized audio sources are placed according to raycasts at the specified angular offsets. Third, the soundscape is synthesized from the virtually placed audio sources. Fourth, this soundscape is played to the user with stereo headphones. Fifth, the user interprets the audio information and uses it to navigate through their environment. As the user moves about their environment the scanning device updates the 3D map within the application, causing an update to the positions of the virtual audio sources and the resulting soundscape.

broad impact by helping to augment the quality of life, vision rehabilitation, and general care for many who are blind or visually impaired.

Principle

Inspired by echolocators who can navigate by making and interpreting the reverberations of short "clicking" sounds [11], this work describes a novel assistive application platform to leverage modern 3D scanning technology on a mobile device to digitally construct a 3D map of a user's surroundings as they move about a space. Within the digital 3D scan of the world, spatialized audio signals are placed to provide the navigator with a real-time 3D stereo audio "soundscape." As the user moves about the world, the soundscape is continuously updated and played back within the navigator's headphones to provide contextual information about the proximity of walls, floors, people, and other features or obstacles. This structure is illustrated in Fig. 1.

To allow for the on-demand creation of virtual environments and for the realistic simulation of spatialized audio, the mobile application was implemented in the Unity game engine [12]. Newer Pro models of the iPhone and iPad were targeted for this research as they are equipped with a built-in, rear-facing LiDAR scanner. With Apple's ARKit [13] (the underlying software development framework for Apple's augmented reality capabilities), depth data from this LiDAR scanner can be used to produce a real-time 3D mesh reconstruction of the user's physical environment. The AR-Foundation [14] plugin was used to interface Unity with this dynamic scene generated by ARKit.

Once a digital reconstruction of the world's geometry is established, the soundscape is created by placing an array of spatialized audio sources within the virtual world. The positions of these audio sources are determined by "raycasting." Raycasting is performed by sending an imaginary line (a ray) into the digital reconstruction of the world in some direction. Any collision of this ray with a part of the world's geometry (e.g., a wall) is detected. Once a collision point is known, an audio source is placed at that location. This process can be repeated for any desired number of rays (audio sources) or initial angles of offset. Figure 2 illustrates this approach.

As a user navigates, each ray is semi-continuously re-cast into the digital reconstruction of the world and the position of the associated audio source is updated. Depending on the world's geometry and the user's perspective, each audio source will have a unique volume intensity (related to how far it is from the user). It will be these amplitude variations that produce the soundscape when spatially combined. The sounds produced by each of the audio sources that compose the soundscape can be individually adjusted (though the optimization of their combination is a limitation yet to be explored). For the current implementation, audio sources have been arbitrarily configured to play a series of pure tones. Irrespective of the specific sounds played, each audio source is spatialized by Unity within the digitized 3D environment. The resulting soundscape is then output to the user via stereo headphones. Apple AirPods Pro were used here as they include modes for active noise cancellation and ambient audio passthrough.

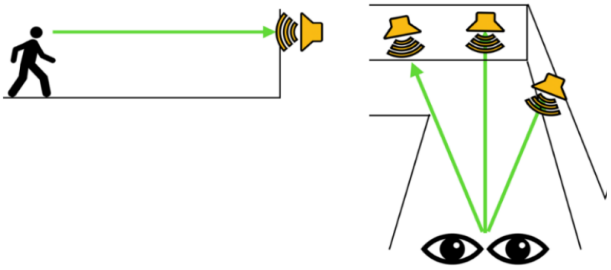

Figure 2. Representation of using raycasting to place audio sources. Left: an audio source is placed where the ray intersects with the wall in front of the user. Right: multiple rays at various angles can be used to place several audio sources; this technique is used to create the spatial audio "soundscape."

Experiments

Testing of the application framework has taken place within ideal virtual environments, digitized environments, and real-world environments. The initial conceptual design and feasibility tests were done within virtual environments, such as ideal corridors or simple mazes. Once a coarse design had been developed, real-world environments were digitized with technologies such as photogrammetry, for more representative testing. Figure 3 shows

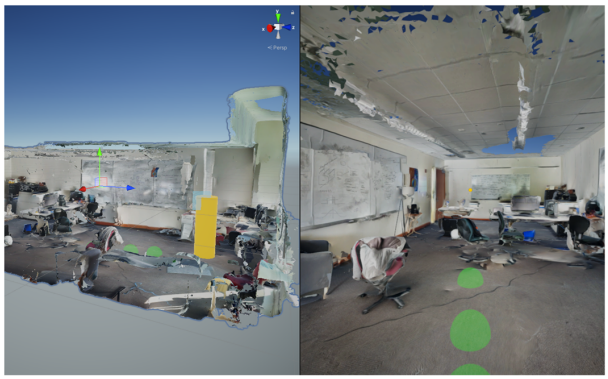

Figure 3. Example use of the application within a digitized real-world environment. Spatial audio sources are represented with green spheres.

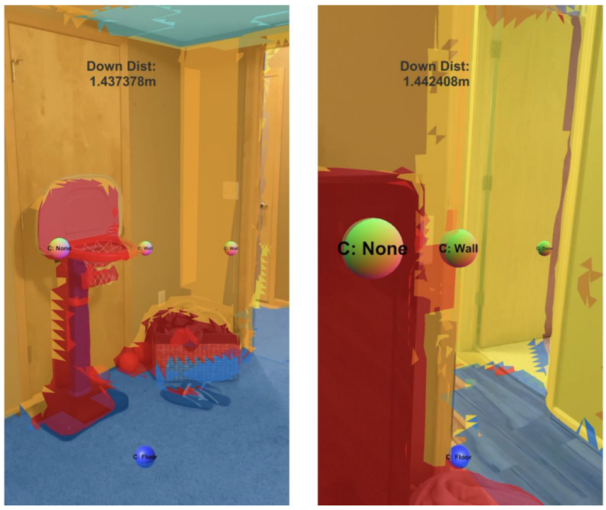

Figure 4. Example real-time use of the application within a real-world environment. The reconstructed 3D scene is represented with color overlays, coded to their classification, and the spatial audio sources are visualized with spheres.

an example of a user training with the soundscape using a digitized environment in the Unity engine and an early sound source configuration. Figure 4 shows an example real-world use of the mobile application within a living room environment. After these feasibility tests, additional experiments were conducted to establish data collection and analysis techniques. In the presented experimental implementation, six audio sources were used in a 't' configuration. The left and right sources were 30° offset from the center source, all three of which currently play a G4 tone. The upper source was 20° offset from center and plays a B4 tone. The other two remaining sources were at 30° and 60° down from the center source and play E4, and C4 tones, respectively. Timestamped positions and orientations from within the reconstructed 3D environment were stored in a database for analysis. These data were collected for the user and for each of the audio sources.

In order for this platform to be tested within a future user study, comparisons between independent test runs must be possible. However, unless the operating device is in precisely the same location at the start of every run, each run's world origin will have a different location and orientation. This is because the origin of each virtual reconstruction produced by ARKit is created afresh for each run. Thus, to compare across test sessions, the data from each run must be registered or aligned to share a common origin. Alignment was enabled by integrating a fiducial marker to the scene in the form of an image-based ARTag. Adding the ARTag allowed a real-world object (in this case a printed calibration target) to be associated with a consistent virtual-world position and orientation. During each test, the marker was placed in a fixed, real-world location and the per-test virtual position and orientation of the marker were recorded in the database. After tests were performed, the stored time series information was retrieved from the database and transformed so that each arbitrarily generated origin was aligned with the fixed, objective, fiducial marker that had been located for each run.

Using these aligned data, velocity and direction of travel could be estimated, enabling more abstract analysis. These more abstract metrics included "seeking," which is defined here as the total head rotation in the plane of travel and "hesitation," a qualitative measure of the degree of confidence demonstrated by a user in their navigation. This hesitation score was introduced in [15] and is calculated as the sum of differences between sequential large peaks and troughs normalized by the maximum traversal speed of the trial (smaller hesitation scores indicate a more consistent rate of traversal). A path trace visualization of an example test run is provided as Fig. 5. In this run, a user ambulates up through a hallway, navigates around an obstacle, turns right, and moves down another hallway. The path trace is shown with a box plotted in the approximate real-world position of the obstacle (a recycling bin). Encoded in the trace color is the normalized traversal speed, red indicating slower traversal and blue indicating faster traversal.

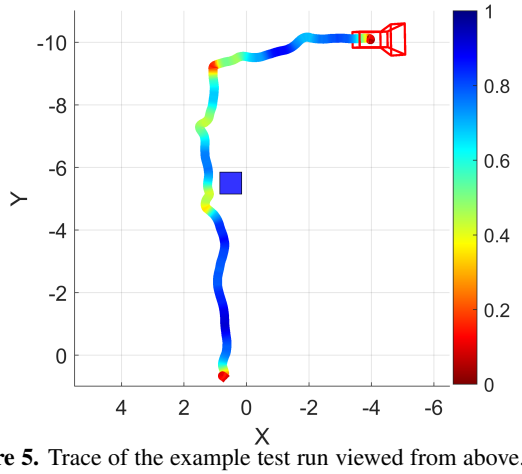

Figure 5. Trace of the example test run viewed from above. The color coding of the trace indicates the normalized traversal speed at that point, red being slower, blue being faster.

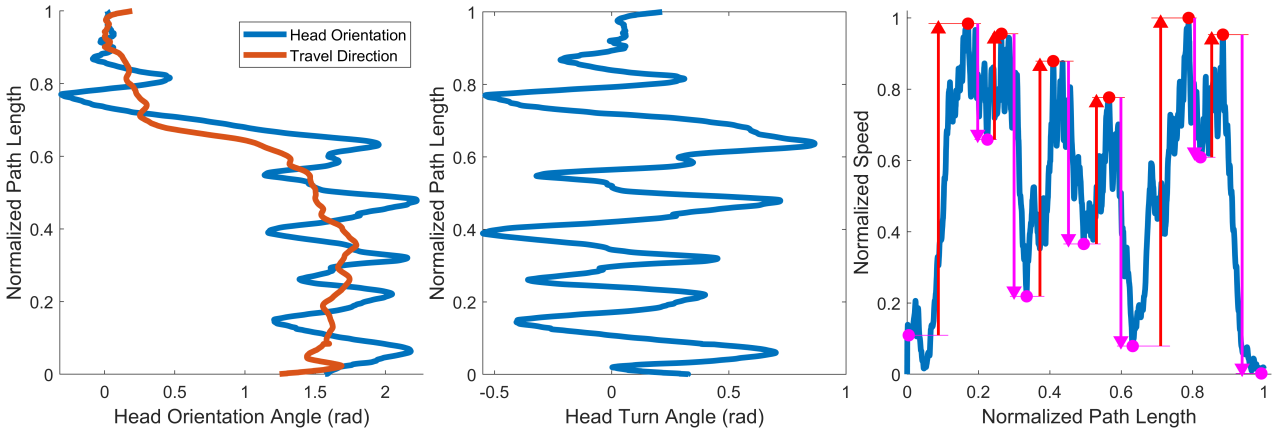

Figure 6. Performance metrics derived from an example test run. (a) Head turn angle and the approximate direction of travel against normalized path length, note the correspondence between (a) and the path trace in Fig. 5, where the sharp decrease in head angle of $\pi/2$ radians matches the right turn in the path trace. (b) Head deviation angle calculated as the difference between the two curves in (a) plotted against normalized path length. These data were used to calculate the total head rotation angle as approximately 1.85π radians. (c) Normalized speed against normalized path length, overlain are the hesitation peaks and troughs and the directional distances between them producing a score of 7.12.

A plot of the head orientation angle and estimated travel direction is provided as Fig. 6(a) (their difference is also shown as head deviation in Fig. 6(b)). A plot of the normalized travel speed is provided as Fig. 6(c), with an overlay of the peaks and troughs used to calculate the hesitation score. Head orientation is plotted against the normalized path length of the test run in Fig. 6(a), along with the direction of travel (which is estimated from the position data). Though all three axes of rotation are recorded to the database, the relevant rotation data for this test run were observed to be mostly contained within the rotation about the Z axis, thus its selection for plotting. The plot of the head deviation angle is presented as Fig. 6(b). Head deviation was calculated as the difference between head orientation and approximate travel direction. Oscillations or "seeking" off-axis from the direction of travel are evident in the plot, cumulatively comprising the final total head rotation value of approximately 1.85π radians (or about 330°). This was calculated as the sum of the absolute value of each prominent extremum (the absolute sum of the maxima and minima of the smoothed head angle deviation data).

The speed of travel at each point was estimated as a difference in position divided by the time differential between the sampling of the points. The peaks and troughs of the velocity curve were extracted; those used to calculate the hesitation score were determined algorithmically. A peak was accepted if three conditions had been met: (1) a prior trough had been established; (2) an increase of more than 25% of the maximum speed was encountered; and (3) if, after (2), another trough more than 25% less than the potential peak was found. The inverse is true for establishing troughs (an initial peak or trough is naively accepted). These peaks and troughs are used to find variations in speed, which are then summed and normalized by the maximum speed, giving the hesitation score. For this example run, the hesitation score was 7.12, qualitatively indicating a modest amount of hesitation.

While a functional prototype of the application has been achieved and is ready to be formally evaluated through a user study, there are a number of elements that are being refined: (1) the selection and arrangement of sound sources; (2) adapting the integration of mesh classification results into the soundscape; and (3) addressing user experience and day-to-day usability for people who are blind or visually impaired.

Summary

This paper has presented the framework for a novel assistive mobile application aimed at improving the safety, efficiency, and independence with which PVI are able to navigate. Navigation assistance is accomplished by building a real-time 3D audio "soundscape" providing users with 3D stereo audio feedback about their surroundings. The application does this by using the LiDAR scanning capabilities of modern mobile devices to build a dynamic virtual 3D map of the user's environment. This mapping is then used to place artificial sound sources at customizable locations throughout the scene. These spatialized sound sources are then played to the user via stereo headphones. A methodology was also presented that will enable the comparative analysis of data collected from different experimental trials of the developed application. Several metrics are currently extracted and computed, providing the foundation for the enabled comparison. These metrics include the 3D position, orientation, and velocity of the user and their derived "seeking" and "hesitation" metrics. Overall, the contributions of this paper enable user studies aimed at optimizing the proposed technology to enhance the safety, efficiency, and independence of PVI navigation.

Funding

The research described here was supported (or supported in part) by the Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service. Award number 1I50RX003002, Tyler Bell, Center for the Prevention and Treatment of Visual Loss, Iowa City VA, Without Compensation Position (WOC).

References

[1] Rohit Varma, Thasarat S Vajaranant, Bruce Burkemper, Shuang Wu, Mina Torres, Chunyi Hsu, Farzana Choudhury, and Roberta McKean-Cowdin. Visual impairment and blindness in adults in the united states: demographic and geographic variations from 2015 to 2050. *JAMA ophthalmology*, 134(7):802–809, 2016.

[2] Ashley D Deemer, Bonnielin K Swenor, Kyoko Fujiwara, James T Deremeik, Nicole C Ross, Danielle M Natale, Chris K Bradley, Frank S Werblin, and Robert W Massof. Preliminary evaluation of two digital image processing strategies for head-mounted magnification for low vision patients. *Translational vision science & technology*, 8(1):23–23, 2019.

[3] Pawel Strumillo, Michal Bujacz, Przemyslaw Baranski, Piotr Skulimowski, Piotr Korbel, Mateusz Owczarek, Krzysztof Tomalczyk, Alin Moldoveanu, and Runar Unnthorsson. Different approaches to aiding blind persons in mobility and navigation in the "Naviton" and "Sound of Vision" projects. In *Mobility of visually impaired people*, pages 435–468. Springer, 2018.

[4] Michal Bujacz, Piotr Skulimowski, and Pawel Strumillo. Naviton—a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired. *Journal of the Audio Engineering Society*, 60(9):696–708, 2012.

[5] Simona Caraiman, Anca Morar, Mateusz Owczarek, Adrian Burlacu, Dariusz Rzeszotarski, Nicolae Botezatu, Paul Herghelegiu, Florica Moldoveanu, Pawel Strumillo, and Alin Moldoveanu. Computer vision for the visually impaired: the sound of vision system. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1480–1489, 2017.

[6] Daniela Massiceti, Stephen Lloyd Hicks, and Joram Jacob van Rheede. Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PloS one*, 13(7):e0199389, 2018.

[7] Caitlin Dodsworth, Liam J Norman, and Lore Thaler. Navigation and perception of spatial layout in virtual echo-acoustic space. *Cognition*, 197:104185, 2020.

[8] Santiago Real and Alvaro Araujo. VES: A Mixed-Reality Development Platform of Navigation Systems for Blind and Visually Impaired. *Sensors*, 21(18):6275, 2021.

[9] L Fialho, Jorge Oliveira, André Filipe, and F Luz. Soundspace VR: spatial navigation using sound in virtual reality. *Virtual Reality*, pages 1–9, 2021.

[10] Kristen Shinohara and Jacob O Wobbrock. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 705–714, 2011.

[11] Andrew J Kolarik, Silvia Cirstea, Shahina Pardhan, and Brian CJ Moore. A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing research*, 310:60–68, 2014.

[12] Unity Technologies. Unity. `https://unity.com/`. v2022.1.3f1.

[13] Apple Inc. ARKit. `https://developer.apple.com/augmented-reality/arkit/`. v6.

[14] Unity Technologies. AR Foundation. `https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@5.0/manual/index.html`. v5.0.

[15] Joram J van Rheede, Iain R Wilson, Rose I Qian, Susan M Downes, Christopher Kennard, and Stephen L Hicks. Improving mobility performance in low vision with a distance-based representation of the visual scene. *Investigative ophthalmology & visual science*, 56(8):4802–4809, 2015.

Author Biography

Broderick S. Schwartz is a Ph.D. candidate in Electrical and Computer Engineering at the University of Iowa. Broderick received his B.S. in Mechanical Engineering from Purdue University in 2019 and his M.S. in Electrical and Computer Engineering from the University of Iowa in 2021. His current research interests include 3D capture and compression.

Prof. Tyler Bell is an Assistant Professor of Electrical and Computer Engineering at the University of Iowa. He leads the Holo Reality Lab and is a faculty member of the Public Digital Arts (PDA) cluster. Tyler received his Ph.D. from Purdue University in 2018. His current research interests include high-quality 3D video communications; high-speed, high-resolution 3D imaging; virtual reality, augmented reality; human computer interaction; and multimedia on mobile devices.