An AR-Guided System for Fast Image-Based Modeling of Indoor Scenes

Daniel Andersen* Voicu Popescu†

Purdue University

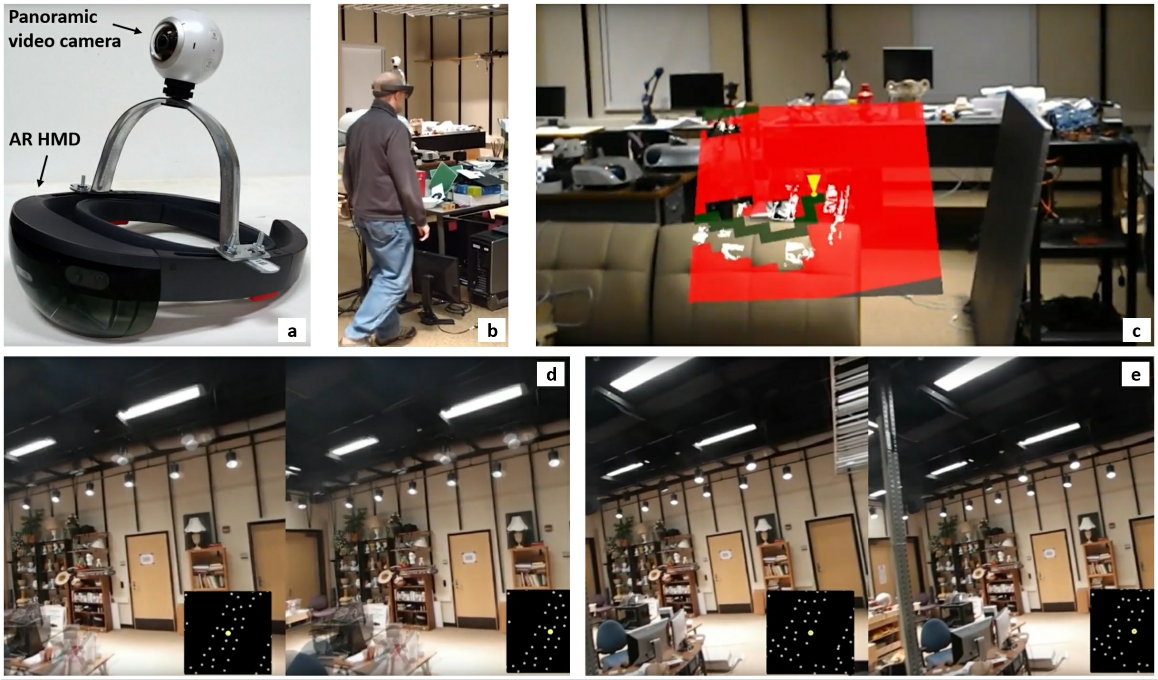

Figure 1: Acquisition device (a), AR-guided acquisition (b), acquisition map visualization (c), guided stereo VR visualization from viewpoint in between acquisition locations (d), and from acquisition location (e).

Abstract

We present a system that enables a novice user to acquire a large indoor scene in minutes as a collection of images that are sufficient for five degrees-of-freedom virtual navigation by image morphing. The user walks through the scene wearing an augmented reality head-mounted display (AR HMD) enhanced with a panoramic video camera. The AR HMD visualizes a 2D grid partitioning of a dynamically generated floor plan, which guides the user to acquire a panorama from each grid cell. The panoramas are registered offline using both AR HMD tracking data and structure-from-motion tools. Feature correspondences are established between neighboring panoramas. The resulting panoramas and correspondences support interactive rendering via image morphing with any view direction and from any viewpoint on the acquisition plane.

Index Terms: Human-centered computing—Mixed / augmented reality; Computing methodologies—Virtual reality; Computing methodologies—Image-based rendering

1 Introduction

Applications such as virtual tourism, real estate advertisement, or cultural heritage preservation involve high-quality interactive rendering of real world scenes. However, efficient photorealistic acquisition of such scenes is a challenging problem. Traditional texture mapped geometric models are difficult to acquire to a level of completeness necessary for high-fidelity rendering. Image-basd modeling and rendering is an alternative approach where the scene is captured as a database of rays, which is queried at runtime to to render from a desired view. Good results are obtained as long as the viewing volume was sampled densely. However, efficient image-based modeling of a large indoor space remains difficult, due to the disadvantages of expensive acquisition devices, of long acquisition times, and of reliance on operator expertise. Without immediate feedback during acquisition, it is difficult for a user to capture the scene reliably from all necessary viewpoints. Returning to the scene after initial acquisition to capture additional viewpoints is often impractical.

We present a system for efficient image-based modeling and rendering of indoor scenes. The acquisition device is an augmented reality head-mounted display (AR HMD) enhanced with a panoramic video camera (Fig. 1a). Under AR guidance, the user acquires a complete and dense set of panoramas of the scene (Fig. 1b). The system displays an acquisition map (Fig. 1c) showing unexplored regions (red), explored regions (green), scene geometry (white), and the user's current location (yellow). The user walks through the scene to acquire a panorama from each accessible cell of a horizontal 2D grid at user head height. Offline processing registers the panoramas using AR HMD tracking data and structure-from-motion. Panoramas are grouped into triplets by proximity and correspondences are found between adjacent panoramas for use in image-based morphing.

We support interactive visualization through a virtual reality (VR) HMD, with five degrees-of-freedom (2D translation and 3D orientation). The desired image is rendered by morphing and blending

*e-mail: andersed@purdue.edu
†e-mail: popescu@purdue.edu

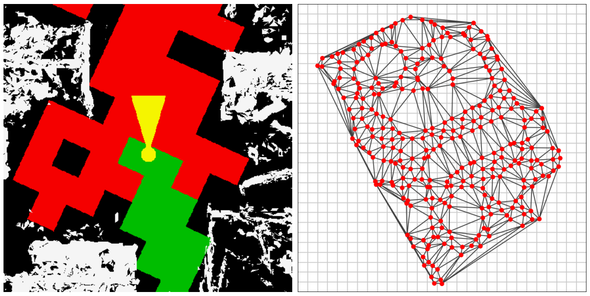

Figure 2: Left: 2D map shown to the user during acquisition. Right: 2D triangulation of panorama positions selected from acquisition path.

the three panoramas that define the triplet containing the current viewpoint. The VR HMD also displays a map with the current user position and the nearby panorama acquisition locations. At viewpoints in between acquisition locations (Fig. 1d), correspondence-rich regions of the scene (e.g. bookshelves, doors, walls, floors) are visualized with high quality, while regions with sparse correspondences exhibit ghosting artifacts. Viewpoints close to an acquisition location (Fig. 1e) converge to the original panorama for artifact-free viewing. Leveraging the high density of the acquisition and the visualization of the acquisition locations, the user achieves a high-quality interactive visualization of the scene, with brief, quality transitions in between acquisition locations, and with photorealistic pan-tilt sequences from viewpoints aligned with the acquisition locations.

2 Technical Approach

The user acquires scene panoramas with interactive guidance from an AR interface. After offline processing, the panoramas are then morphed to support interactive VR navigation of the scene.

2.1 Acquisition Device and AR Interface

The user acquires the scene while wearing an AR HMD enhanced with a panoramic video camera (Fig. 1a). An AR interface guides the user towards a fast, dense, and complete acquisition of a uniform 2D grid of panoramas on a horizontal plane. This grid appears as a map floating in front of the user (Fig. 1c, Fig. 2 left). The map shows unexplored regions (red), inaccessible regions due to floor obstacles (white and black), explored regions (green), and the user's current viewpoint (yellow). Accessible regions are computed using A* navigation against the AR HMD's estimated geometry of the scene (acquired through active depth sensing).

2.2 Offline Processing

Acquisition results in a video sequence of panoramas and a sequence of estimated AR HMD poses. The sequence of poses is partitioned into a 2D grid matching the cell size shown during acquisition. The panorama matching the pose closest to each traversed cell's center is selected. We then generate a 2D Delaunay triangulation from the horizontal positions of each selected frame, resulting in a set of triplets whose vertices are its surrounding panoramas (Fig. 2, right).

Because the AR HMD's estimated poses are imprecise, we do sparse 3D reconstruction by conventional structure-from-motion, with panoramic cubemap faces as input [5]. We perform an additional custom bundle adjustment routine to ensure all faces of a single cubemap remain precisely aligned with each other. For each triplet in the triangulation, we record the subset of sparse 3D points that were visible to at least one panorama in the triplet. The 3D points are combined with those vertices in the AR HMD's rough geometric model that are visible to viewpoints within the triplet.

We triangulate each triplet's observed 3D points such that mesh topology is consistent when projected onto a unit sphere located at any point within the triplet. We follow the method of Zhao et al. [6] to iteratively filter any topologically inconsistent 3D points.

2.3 Interactive Visualization

Our visualization method extends the spherical image morphing method of Zhao et al. [6] and Kawai et al. [2] by allowing five-degree-of-freedom motion of a virtual camera through the 2D scanned region across multiple sets of adjacent panoramas.

The system determines the triplet surrounding the user's current virtual camera position. The triplet's correspondence points are rendered as a spherical texture surrounding the virtual camera. Each pixel of the spherical texture represents a ray originating from the virtual camera position and encodes where that ray collided with the triangulated 3D points. These values act as lookup values into the triplet's panorama images to find the texture coordinates to achieve a smooth transition between panoramas in the triplet.

The user can view the scene with a VR HMD to freely move the virtual camera (Fig. 1d and Fig. 1e). Because visualization quality is highest when the virtual camera matches an original acquisition point, we provide a top-down map of the acquisition grid while in VR (Fig. 1d and Fig. 1e, bottom right corner of frames), which allows the user to orient themselves to ideal viewing locations.

3 Results and Discussion

Our AR acquisition device uses a Microsoft HoloLens [3] AR HMD coupled with a Samsung Gear 360 panoramic camera [4]. The camera captures panoramic frames at a resolution of 3,840 x 1,920 pixels. The VR visualization runs on the HTC Vive HMD [1].

First-time users of our systems with general AR/VR experience could acquire a complex indoor scene (10m x 15m, grid cell size = 0.5m) in 6min30s on average. Both the AR interface for acquisition and the VR morphing visualization run in real time. Offline processing time is about 2 hours, depending on the number of panoramas.

Fig. 1 shows visualization frames obtained for a complex, cluttered indoor environment of approximately 10m x 15m. Acquisition time was 5min14s. A total of 304 panoramas were selected (grid cell size: 0.5m). The panoramas were grouped into 589 triplets, with an average of 1,998 correspondences per triplet, which are sufficient for quality interactive visualization of the scene. Transitions are brief, clear, and not disorienting, and the provided guidance enables the user to align the viewpoint with acquisition locations.

3.1 Conclusion and Future Work

Our system enables fast image-based modeling and rendering of indoor spaces. Using an AR interface to guide the user towards a complete a dense acquisition, novice users can acquire a large indoor scene in minutes, suitable for image-based morphing in VR.

Some distortion is visible when the virtual camera is far from the closest acquisition location. Because correspondence topology within triplets is enforced, local changes in occlusion will result in ghosting. Future work will use the AR HMD's rough model of scene geometry to dynamically adjust grid cell size during acquisition.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant DGE-1333468. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the NSF.

References

[1] HTC. VIVE, 2017.
[2] N. Kawai, C. Audras, S. Tabata, and T. Matsubara. Panorama image interpolation for real-time walkthrough. In *ACM SIGGRAPH 2016 Posters*, SIGGRAPH '16, pp. 33:1–33:2. ACM, New York, NY, USA, 2016. doi: 10.1145/2945078.2945111
[3] Microsoft. Microsoft HoloLens, 2017.
[4] Samsung. Gear 360 Camera, 2017.
[5] C. Wu. VisualSFM: A visual structure from motion system, 2011.
[6] Q. Zhao, L. Wan, W. Feng, J. Zhang, and T. T. Wong. Cube2Video: Navigate between cubic panoramas in real-time. *IEEE Transactions on Multimedia*, 15(8):1745–1754, Dec 2013. doi: 10.1109/TMM.2013.2280249