Indoor Planning Using Diminished and Augmented Reality

WERNER BAILER and HANNES FASSOLD, JOANNEUM RESEARCH, Austria
VASILEIOS GKITSAS, PETROS DRAKOULIS, and DIMITRIOS ZARPALAS, Information Technologies Institute / Centre for Research and Technology, Greece
ROBERT HUEMER, Roomle GmbH, Austria

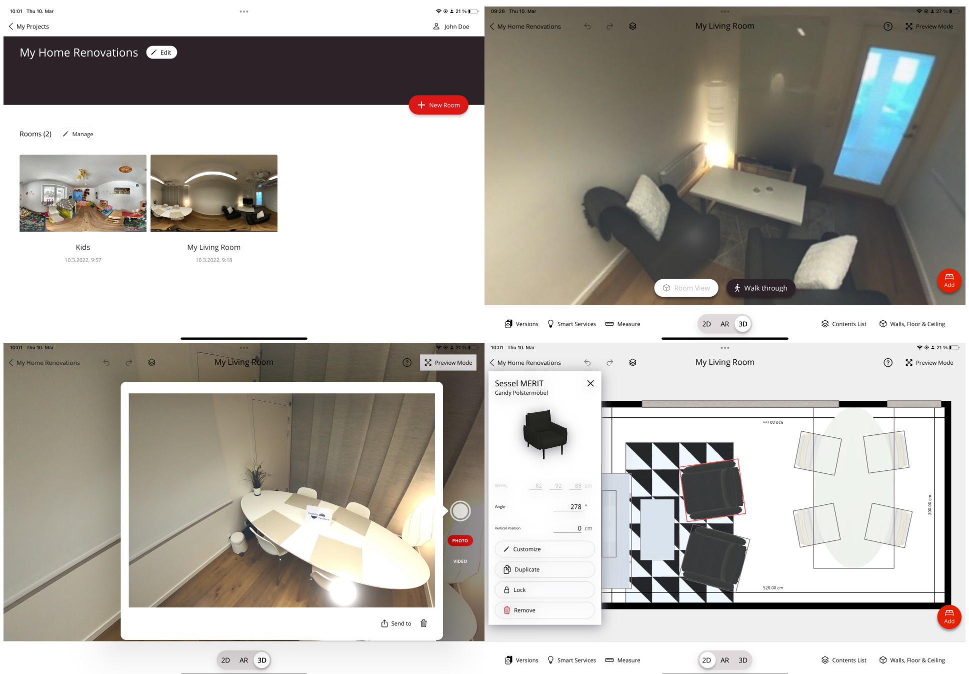

Fig. 1. Screenshots of the prototype application for XR-supported indoor planning: project overview, AR walkthrough, snapshot of views and floorplan (top left to bottom right).

Virtual and augmented reality has the potential to improve collaboration between professionals and customers in indoor planning tasks like interior design or refurnishing. Our work addresses two major issues of current AR planning apps: the required technical skills to create or import room plans, and the overlap between existing and overlaid items. We propose automation of the scene creation from a single panorama using AI-based services and the use of diminished reality (DR) to conceal objects before overlaying new ones.

Additional Key Words and Phrases: augmented reality, indoor planning, scene understanding

1 INTRODUCTION

Applications like interior design, furniture retailing and renovation can be a challenging process, requiring bidirectional communication between professionals and customers or future users during the planning and

IMX '22 - XRWALC, 2022, Portugal

design phase. Making this communication process effective saves costs, avoids later modifications, and results in providing tailored solutions and higher customer satisfaction. Experts usually express their ideas in traditional 2D drawings produced by computer aided design (CAD) software, making it difficult for the consumers to comprehend them. Augmented reality (AR) is an emerging technology with potential to facilitate this communication process, as it allows superimposing virtual 3D objects within the real environment. In this way, AR serves as a medium between digitized concepts and the real scene, facilitating effective and efficient communication and feedback between its users, improving the iterative design process. Complementing AR with virtual reality (VR) enables visualizing planning both on and off site, i.e., consumers can view planning overlaid over real environments, while the virtual scene can be edited by professionals in the office.

A number of AR indoor planning apps already exist (for pointers to many of them see [9]).

However, we can observe two main limitations of the existing apps: First, creation of the room layout needs to be done manually in most apps, or via importing a CAD model. This requires some IT and technical skills, and may be an obstacle for users. Second, indoor design does in many cases not start from scratch in an empty room, but makes changes to an existing room. When viewing changes on site, the realism of AR is severely degraded when the overlaid 3D objects added to the scene clash with real objects. AR of additive nature, with causes a practical problem when working in occupied and filled indoor scenes as is the case for AR home design applications [11].

The first issue can be addressed by *automation*. In order to enable users to create scenes without specific skills, we aim to create the scene from capturing panoramic images. Recent advances in artificial intelligence (AI)-based visual scene understanding enable this automation for constrained environments (such as private or office indoor scenes). These technologies enable detecting the layout of the captured room as well as present objects, and their boundaries and dimensions. Merging the roles of *user and creator* also blurs the line between authoring and consuming an AR/VR scene, which are currently often done in specific tools. The second issue can be addressed by *diminished reality (DR)* technologies, which enable visually concealing real objects. In current AR authoring and presentation solutions, the support for DR is still very limited. As the content behind the object to be removed is unknown, DR needs to hallucinate content, typically referred to as infilling or inpainting [6]. Pioneering work in the DR domain was presented by [3], where a patch-based image inpainting method was developed. Follow up work [4] moved beyond image based diminishing and transitioned towards respecting scene geometry by exploiting SLAM-based localization. More recently, an inpainting method for non-planar scenes was developed [5] that considered both color and depth information. Still, in both cases, manual selection of the region to be removed in the image domain was required. To allow for easier selection of the object to be diminished in indoor scenes for interior design, [7] used a manually positioned and scaled volume to enclose the object of interest. In addition, the floor plane was identified by inserting a marker into the scene.

In order to diminish real objects with minimal user interaction, existing objects in the scene must be segmented, allowing to select the region to be removed with a single selection, and the background must be reconstructed with methods such as inpainting. Further details from our analysis of user requirements can be found in [9].

An overview of the proposed processing workflow is shown in Figure 2. Starting from a panorama of the room, the system first estimates the room corners and performs instance segmentation of the object in the scene. Using the mask from the segmentation, object regions can be inpainted in order to prepare the content for diminishing them. The region for the diminished object is blended with the panorama and fed into the augmentation pipeline. Details about the AI-based methods for the components can be found in [1]. For visualization using AR on-site, the augmentation is performed using the toolkit of the target platform to overlay the inpainted region and the rendered 3D objects (currently, the application is implemented for iOS using ARKit). For visualization using VR off-site, the virtual parts of the scene are rendered, with the captured panorama (with inpainted regions for diminished objects, if applicable) as a backdrop of the scene. This enables similar views and interaction capabilities for both professionals and consumers, whether they are on or off-site.

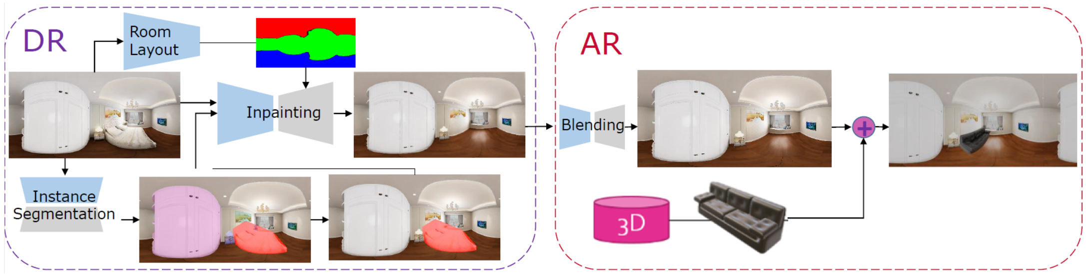

Fig. 2. Proposed processing workflow.

The rest of this paper is organized as follows. Section 2 provides an overview of the system design, architecture and the backend services. Section 3 presents the app prototype, and Section 4 discusses the approach for evaluation. Section 5 concludes the paper and outlines future work.

2 SYSTEM OVERVIEW

The system design follows a layered approach, clearly separating the frontend applications, data handling and backend processing services. The main design principles are distributed, service-oriented and event-driven. The architecture has been designed under the assumption of a fully distributed system, where not only the user applications but also the data stores and components may be stored at different locations and/or be provided by different organizations. This may apply to AI services that need to run on public cloud services to enable high scalability, object catalogues provided by B2B customers or user data stored with a B2B customer. Further, this architecture decouples the user-facing application layer from the functionality implementing layer (AI services), facilitating the integration of new and emerging AR/VR devices (e.g. the Microsoft HoloLens 2.0), which is important in the quickly evolving landscape of interactive technologies.

Most AI and processing functionalities are provided as services, each of them performing one clearly specified function on a single or multiple data item(s), the latter focusing on multi-modal inputs. The services are provided as REST services with a well-defined API as well as input/output metadata structures. The system is controlled in an event-driven paradigm, where events are generated by added or modified data items (e.g. panorama being captured) or explicit user interactions. Events are also used to trigger processing by the services and notify the availability of results. In order to decouple the services from the specific data store and event mechanism, middleware components serve as connectors, feeding the REST interface with the required inputs and relaying the results back to the data layer.

2.1 Architecture

The system architecture is organized into four layers: applications, data, middleware and services. The *applications* layer represents the user-facing application of the system, which is realized as a mobile application used on-site (i.e. in the room being planned) or off-site (e.g., in the room next door, in the office, at a sales outlet etc.). The *data* layer represents all data stores used by the system, which includes a NoSQL database storing relevant metadata of the system, as well as links to assets (e.g. images, point clouds and 3D models) stored on cloud storage. The *middleware* layer acts as an abstraction barrier, decoupling the services from specific implementation details realized in data and application layers. The *services* layer represents the AI tools for (mostly non-real time) processing of data.

2.2 Backend services

The data of captured scenes is being stored in the cloud in two different ways: World knowledge, i.e. data describing the scene, its objects and the authoring decisions, are stored in a NoSQL database. In particular, Google Firebase[1] is used as the platform which enables access from both cloud-based services and mobile apps as well. Assets such as images, 3D models and their textures are stored in cloud object storage.

The AI services provide modular capabilities performing a particular task, typically processing a media asset or a set of metadata items, and output media assets or metadata items that can be leveraged by subsequent AI services or directly used by the application layer. The AI services are implemented in using TorchServe[2] to provide a standard REST interface for inference. *Layout estimation* takes a panorama as input and provides metadata describing the room layout (sparse/planar scene geometry). *Instance segmentation* takes a view or a panorama as input and provides metadata with object bounding polygons and class labels. *Depth estimation* takes a single or multiple overlapping panoramas as input and provides an estimated depth map (dense scene geometry) and the reconstructed 3D model of the scene involved. *Panorama inpainting* takes a view or a panorama, as well as object polygons or masks as input and provides a set of image patches or a newly composited panorama for replacing each of the objects. Further, it provides the predicted dense layout along with the empty, non-furnished, version of the input panorama. *Scene localisation* uses a panoramic image and the instance segmentation and depth estimation results to propose a set of image patches in the scene which are suitable as image anchors, which the AR app uses or registering the AR scene. In addition, services for estimating a scene graph and proposing alternative furniture arrangements are under development. More details about the AI services can be found in a project report [8].

3 APP PROTOTYPE

The work on an application prototype was preceded by collaborative workshops used to gather user requirements and developing UI/UX designs. This laid the groundwork for the implementation of an iterative prototype. The prototype is currently available as an iOS application. Some screenshots are shown in Figure 1.

After registering an account, the user can create projects, which can contain multiple rooms. A project can be shared with other users (family, friends, business partners, . . .). Collaboration of multiple users is possible by adding editing permissions. View-only rights can also be applied on a per-user basis. In a project, users can create rooms by uploading a captured 360° panorama of a room. The data of a room is analysed and pre-calculated in the middleware to allow offline use once the data is completed and transferred to the client application.

If a user alters the room (e.g., removes/adds objects) all changes are immediately visible and automatically saved. Calculated bounding boxes are shown around this objects to indicate the possibility of interaction and manipulation. Real objects can be selected and removed, by replacing the texture of the object with the pre-calculated inpainted image provided by the panorama inpainting service. Additional virtual content can be added to the scene from the object catalog, which consists of a large variety of products of known brands as well as generic resize-able objects. Multiple views are available, and direct manipulation is possible in every view. The 2D view gives an orthographic overview of the room, with the automatically constructed layout of the room and all detected objects found by the segmentation service. The AR view allows the fusion of the generated 3D scene with the real world, but it is only available on location, and we currently support 3DOF only, which limits the movement a rather close position to the place the panorama image was originally captured. There is also a VR view available, which consists of a bird's eye view and a walk-around mode. With the touch of one button it is possible to switch to *preview* mode, in which no manipulation is possible. In this mode, all bounding boxes and

[1] https://firebase.google.com
[2] https://pytorch.org/serve/

authoring controls are hidden. This allows for viewing the finished room without obstruction. It is possible to create snapshots and video recordings of the scene, which can be shared with other users.

AI-based layout suggestions are available and grouped together under *smart services*. It is also possible to try out alternative furniture layouts of all detected objects. Re-furnishing a room with proposed alternative furniture, based on criteria like style or price, is another option. Complementing a room with additional, missing furniture is available depending on the analysis of detected objects in the context of the room category. For convenience, an option to completely empty a room is also available, which applies diminishing to every object in the room. Furthermore it is possible to save intermediate versions of the scene, for instance saving different design ideas. A measure tool for measuring the distance between two point in the scene complements the functionality.

4 EVALUATION

The evaluation methodology includes both the objective evaluation of the AI services on appropriate datasets for the respective tasks, as well as subjective evaluation of the app. This includes task-based evaluation on a set of planning tasks, for which we consider the apps Roomle[3], Planner5D[4] and MyTy[5] as baselines, as together they span a similar range of features (lacking the DR support). Preliminary evaluations have been performed with different version of the prototypes and the final evaluation takes place in summer 2022 (the results are published in [10]). The preliminary results show that the tool improves efficiency of users in performing refurnishing tasks, and that the user satisfaction (measured using SUS [2]) is higher than for the baseline tools. In addition, evaluation using walkthroughs with professionals are performed. In a separate set of subjective experiments, the benefit of DR-enhanced AR over AR has been assessed, and the results show that user prefer the DR enhancement over placing objects over real ones (despite inpainting artifacts in some of the scenes) [1].

5 CONCLUSION AND OUTLOOK

We have presented a system for indoor planning, enabling the creation of AR/VR scenes from a single panorama, and supporting diminished reality (DR) for avoiding that inserted objects clash with real objects in AR. The system supports viewing the scene on- (in AR) and off- (in VR) site, thus supporting collaboration between professionals and consumers. In order to enable 6DOF movement in the AR experience, we are working on live DR support by updating the segmentation and inpainting continuously while the viewer through the scene.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 951900 ATLANTIS ("Authoring tool for indoor augmented and diminished reality experiences").

REFERENCES

[1] Georgios Albanis, Vasileios Gkitsas, Nikolaos Zioulis, Stefanie Onsori-Wechtitsch, Richard Whitehand, Per Ström, and Dimitrios Zarapalas. 2022. An AI-based system offering automatic DR-enhanced AR for indoor scenes. In *International Conference on Artificial Intelligence and Virtual Reality*.

[2] John Brooke. 1996. SUS: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996).

[3] Jan Herling and Wolfgang Broll. 2010. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 207–212.

[4] Norihiko Kawai, Tomokazu Sato, and Naokazu Yokoya. 2015. Diminished reality based on image inpainting considering background geometry. *IEEE transactions on visualization and computer graphics* 22, 3 (2015), 1236–1247.

[3] https://www.roomle.com

[4] https://planner5d.com

[5] https://myty.app

[5] Shohei Mori, Okan Erat, Wolfgang Broll, Hideo Saito, Dieter Schmalstieg, and Denis Kalkofen. 2020. InpaintFusion: incremental RGB-D inpainting for 3D scenes. *IEEE Transactions on Visualization and Computer Graphics* 26, 10 (2020), 2994–3007.

[6] Shohei Mori, Sei Ikeda, and Hideo Saito. 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications* 9, 1 (2017), 1–14.

[7] Sanni Siltanen. 2017. Diminished reality for augmented reality interior design. *The Visual Computer* 33, 2 (2017), 193–208.

[8] Vladimiros Sterzentsenko, Georgios Albanis, Nikolaos Zioulis, Vasileios Gkitsas, Antonis Karakottas, Petros Drakoulis, Pachalina Medentzidou, Alexandros Doumanoglou, Dimitrios Zarpalas, Werner Bailer, and et al. 2022. D3.2 AI-based Indoor Scene Understanding v2. https://doi.org/10.5281/zenodo.6581835

[9] Richard Whitehand, Georgios Albanis, Nikolaos Zioulis, Werner Bailer, Dimitrios Zarpalas, and Petros Daras. 2021. Towards User Generated AR Experiences: Enable consumers to generate their own AR experiences for planning indoor spaces. In *ACM International Conference on Interactive Media Experiences*. 299–304.

[10] Richard Whitehand, Per Ström, Brita Piovesan, Robert Huemer, Vasileios Gkitsas, Petros Drakoulis, Antonis Karakottas, Alexandros Doumanoglou, Dimitrios Zarpalas, Werner Bailer, and Stefanie Onsori-Wechtitsch. 2022. D5.4 Evaluation report v2. https://doi.org/10.5281/zenodo.6581892

[11] Kevin Wong, Salma Jiddi, Yacine Alami, Philip Guindi, Brian Totty, Qing Guo, Michael Otrada, and Paul Gauthier. 2020. Exploiting ARKit Depth Maps for Mixed Reality Home Design.