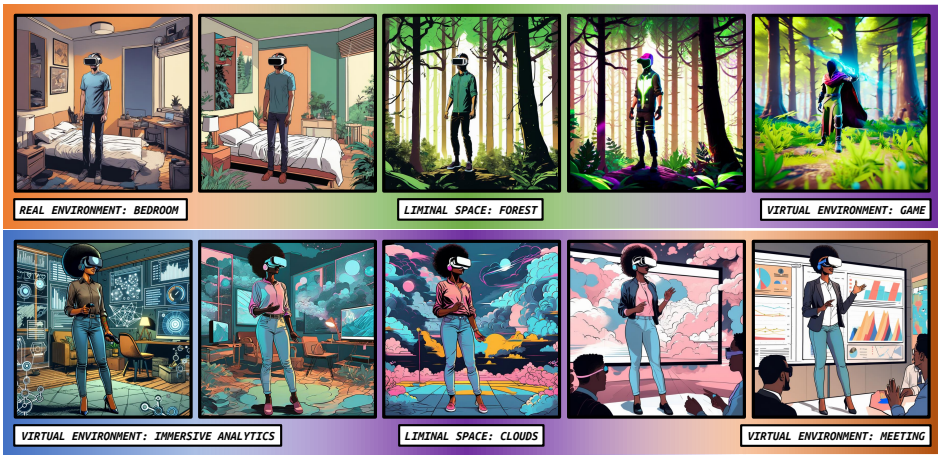
# rlty2rlty: Transitioning Between Realities with Generative AI

Matt Gottsacker*
University of Central Florida

Gerd Bruder†
University of Central Florida

Gregory F. Welch‡
University of Central Florida




(a) A man starts a virtual reality game. He sees his bedroom (real environment) transition into a forest (generated liminal space) that visually foreshadows the game he then plays (virtual environment).

(b) A data analyst reviews her work and prepares some visualizations using immersive analytics tools (virtual environment) for an upcoming meeting. Her environment transforms into a peaceful cloud terrace (generated liminal space) as she briefly meditates and collects her thoughts before joining a meeting (virtual environment) where she presents her findings.

Figure 1: Brief user stories illustrating transitions between realities using our generative system. Users' environments gradually change between their start and end environments, passing through liminal spaces generated for (a) narrative foreshadowing purposes and (b) an intentional meditation break between tasks. This figure's transition images themselves were produced with the generative transition techniques described in this abstract.

## Abstract

We present a system for visually transitioning a mixed reality (MR) user between two arbitrary realities (e.g., between two virtual worlds or between the real environment and a virtual world). The system uses artificial intelligence (AI) to generate a 360° video that transforms the user's starting environment to another environment, passing through a liminal space that could help them relax between tasks or prepare them for the ending environment. The video can then be viewed on an MR headset.

**Index Terms:** Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

## 1 Introduction & Background

Transitioning between tasks and environments is a routine part of using mixed reality (MR) systems, and achieving efficient transitions that make for a fluid cognitive experience is challenging. Practically speaking, compared to abrupt task switches, strategic breaks between tasks can enhance vigor and performance while reducing fatigue [1]. Moreover, from an aesthetic perspective, a gradual transition into a narrative-based immersive experience can improve the audience's experience [9,6]. However, creating good MR transitions is challenging and demands substantial creative and technical effort. In this demo, we present an approach for generating MR transitions between two distinct environments using artificial intelligence (AI) to serve both practical and aesthetic needs.

Researchers have pointed out benefits of smooth transitions and interactions between MR environments, including improving the user experience [9, 6, 2] and increasing presence [13, 3]. And, there has been substantial work on designing methods that support smooth MR transitions. Pointecker et al. [11] designed and evaluated techniques for transitioning between environments, finding that users preferred a fading transition when looking for efficiency and simplicity, and they preferred a more stimulating portal transition between substantively different environments. Kitson et al. [6] found that designing a multisensory transition to prepare users physically, mentally, socially, and environmentally for an awe-inspiring MR experience supported participants' profound emotional experiences. In an interview study, Knibbe et al. [7] found that when exiting virtual environments, users experienced disparities in spatial awareness and orientation, control, sociality, sensory stimuli, and mental presence; each of these presents an opportunity for easing the user out of the environment with a well-designed transition. Our system contributes to this body of research by applying concepts from related work in an automated and general fashion, with the hope that future work can use, refine, and extend our approach to create better MR transitions.

## 2 Reality Transition System

Our functional research prototype visually morphs between two different equirectangular 360° images to provide a continuous aesthetic experience when in transition. Using 360° images for transitions is beneficial because they are simple to capture with a smartphone or 360° camera, no knowledge of the environment's 3D geometry is required, and they can be used with pre-trained AI image models. Additionally, if 3D scene geometry of an environment has already been captured (e.g., as part of scene understanding algorithms or in a 3D environment), it is possible to programmatically convert the environment into a 360° image.

Our fully automated proof-of-concept system uses several AI techniques for analyzing and generating images to produce a visual transition video between any two environments, passing through a liminal space described by a user. The system runs on an MSI Raider GE76 laptop with an NVIDIA RTX 3080 Ti GPU with 16

*e-mail: matthew.gottsacker@ucf.edu
†e-mail: bruder@ucf.edu
‡e-mail: welch@ucf.edu

GB of GDDR6 video memory, an Intel Core i9 CPU, and 32 GB of DDR5 memory. Code[1] and demo videos[2] are available.

**User-Specified Liminal Space** Using a text prompt as guidance, our system can generate a *liminal space* between the starting environment (SE) and ending environment (EE) images. We define "liminal space" to refer to an environment that is in the process of transforming between two different realities, e.g., between the real world and a virtual world or between two virtual worlds. Our process to create the liminal space takes advantage of several features of generative image models to go beyond basic video motion interpolation and produce a smooth transition that is more visually interesting while also being guided by the spatial structure of the input images, which can ameliorate spatial disorientation when transitioning. The liminal space can be utilized for a variety of purposes. For example, it can provide the user with a restorative meditation break between tasks and environments. And, when generated based on a prompt by the environment designer, the liminal space can work in a narrative sense by giving the designer an opportunity to visually foreshadow the user's target environment, which can lead to a better user experience [6]. Alternatively, the user can describe a liminal space they would like to see, which gives them some creative agency in the experience and can also improve their experience [6].

**Image Generation** Our system uses Stable Diffusion XL v1.0's image-to-image synthesis mode [10] to generate images based on the input 360° images of the SE and EE, as well as the user text prompt describing a liminal space. The AI system also uses a *denoising strength* parameter to control the influence of the text versus the image: higher values produce images visually closer to the text prompt than the base image. Our system uses two techniques to ensure that the generated images conform to the visual structure of equirectangular 360° images and display properly on MR HWDs. First, the system uses a low-rank adaptation (LoRA) layer trained on equirectangular 360° images to provide the general visual structure [4]. The system also creates a depth mask for the input images using the MiDaS monocular depth estimator [12], which is then used by a T2I-Adapter [8] to provide additional spatial conditioning to the generated image.

**Transition Construction** The system generates the transition in three phases. In the *start phase*, it generates a series of images based on the SE image that gradually increase the denoising strength while fixing the weight of the T2I-Adapter control for the SE image's depth mask to its maximum and setting the weight of the EE image's depth mask adapter to zero. In the *liminal phase*, it generates a series of images using Stable Diffusion's text-to-image mode (i.e., without basing the images on either the SE or EE image) that begins with the SE image's depth mask adapter maxed out and the EE image's adapter at zero. The generations in this series increase the SE image's depth mask adapter and decrease the EE image's adapter by the same amount each step. In the *end phase*, it inverts the *start phase* using the EE input image as its base, with maximizing the EE image's depth mask adapter while zeroing out the SE input image's adapter. Once all images are generated, the system smooths the transition by inserting additional frames using real-time intermediate flow estimation (RIFE), a model designed to interpolate motion between consecutive images [5]. All frames are then assembled in order and displayed on an MR HWD.

The visual effect in the *start phase* is that images gradually become more like the AI's most intense application of the text prompt while maintaining the depth and visual structure of the SE. In the *liminal phase*, the images are all intense applications of the text prompt while gradually switching the structure of the environment from the SE to match the EE image. The *end phase* is an inverse of the *start phase*, where the environment morphs from the liminal space to the EE with visual structure guidance from the EE.

**Future Work** First, we intend to evaluate how these transition techniques affect users' presence, attention, and task performance at different points in a transition between two environments. From a system perspective, future work could explore ways to allow more user interaction, e.g., to support pausing the transition while the system generates additional scenes in the event the user wants to spend more time exploring an interesting liminal space. Additionally, users may want to control factors of the transition, e.g., how fast different phases occur, through the text prompt. Last, future work could use AI to classify objects in each image to generate transitions based on scene semantics, or transition scene meshes and textures.

## 3 Conclusion

This demo abstract presented an AI-based system that generates a 360° video that visually transitions a user between two arbitrary environments (real or virtual). We described the system's components and how it might be usefully deployed to improve the user's experience when switching realities.


## Acknowledgments

This material includes work supported in part by the Office of Naval Research under Award Numbers N00014-21-1-2578 and N00014-21-1-2882 (Dr. Peter Squire, Code 34), and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).


## References


[1] P. Albulescu, I. Macsinga, A. Rusu, C. Sulea, A. Bodnaru, and B. T. Tulbure. "Give me a break!" a systematic review and meta-analysis on the efficacy of micro-breaks for increasing well-being and performance. *Plos one*, 17(8):e0272460, 2022. 1

[2] M. Gottsacker. Balancing realities by improving cross-reality interactions. *IEEE VR*, pp. 944–945, 2022. 1

[3] M. Gottsacker, N. Norouzi, K. Kim, G. Bruder, and G. Welch. Diegetic representations for seamless cross-reality interruptions. *IEEE ISMAR*, pp. 310–319, 2021. 1

[4] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. LoRA: Low-rank adaptation of large language models. *preprint arXiv:2106.09685*, 2021. 2

[5] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou. Real-time intermediate flow estimation for video frame interpolation. In *European Conference on Computer Vision*, pp. 624–642. Springer, 2022. 2

[6] A. Kitson, E. R. Stepanova, I. A. Aguilar, N. Wainwright, and B. E. Riecke. Designing mind (set) and setting for profound emotional experiences in virtual reality. *ACM DIS*, pp. 655–668, 2020. 1, 2

[7] J. Knibbe, J. Schjerlund, M. Petraeus, and K. Hornbæk. The dream is collapsing: the experience of exiting VR. *ACM CHI*, pp. 1–13, 2018. 1

[8] C. Mou, X. Wang, L. Xie, J. Zhang, Z. Qi, Y. Shan, and X. Qie. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. *preprint arXiv:2302.08453*, 2023. 2

[9] R. Pausch, J. Snoddy, R. Taylor, S. Watson, and E. Haseltine. Disney's Aladdin: first steps toward storytelling in virtual reality. *ACM SIGGRAPH*, pp. 193–203, 1996. 1

[10] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. *preprint arXiv:2307.01952*, 2023. 2

[11] F. Pointecker, J. Friedl, D. Schwajda, H.-C. Jetter, and C. Anthes. Bridging the gap across realities: Visual transitions between virtual and augmented reality. *IEEE ISMAR*, pp. 827–836, 2022. 1

[12] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE TPAMI*, 44(3):1623–1637, 2020. 2

[13] F. Steinicke, G. Bruder, K. Hinrichs, A. Steed, and A. L. Gerlach. Does a gradual transition to the virtual world increase presence? In *2009 IEEE Virtual Reality Conference*, pp. 203–210. IEEE, 2009. 1


[1] https://github.com/mott-lab/rlty2rlty

[2] https://www.youtube.com/watch?v=u4CyvdE3Y3g