# Haptic and Auditory Feedback on Immersive Media in Virtual Reality

Vanessa Pfeiffer
vanessa.pfeiffer<at>stud-mail.uni-wuerzburg.de
University Würzburg
Würzburg, Germany

Sebastian von Mammen
sebastian.von.mammen<at>uni-wuerzburg.de
University Würzburg
Würzburg, Germany

Daniel Pohl
daniel.pohl<at>immerVR.com
immerVR GmbH
Hessdorf, Germany

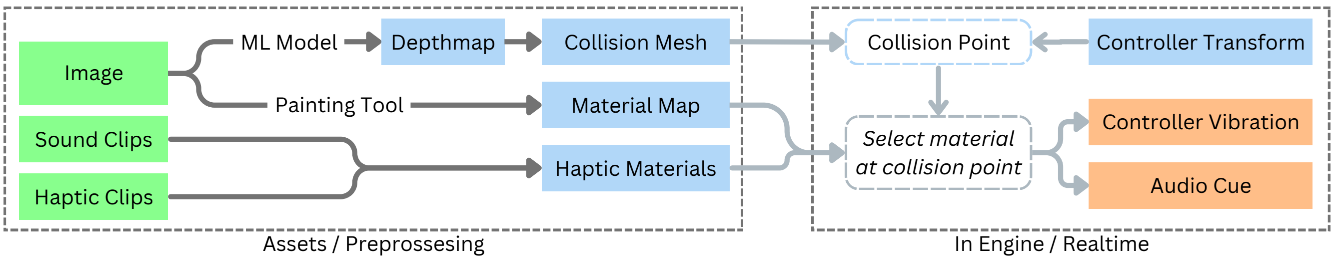



**Figure 1: Using an image as input, we derive a depth map to create a collision mesh. From the image, we design a material map that corresponds to certain sound clips and haptic controller clips that form a haptic material. When a VR controller triggers a collision, we select the haptic material, generate controller vibration, and play back the material's associated sound clip.**

## ABSTRACT

In Virtual Reality (VR), visual and auditory sensations are effectively leveraged to create immersive experiences. However, touch is significantly underutilized in immersive media. We enhance the VR image viewing experience by integrating haptic and auditory feedback into 3D environments constructed from immersive media. We address the challenges of utilizing depth maps from various image formats to create intractable environments. The VR experience is enhanced using vibrohaptic feedback and audio cues triggered by controller collisions with haptic materials.

## KEYWORDS

Haptic Feedback, Auditory Feedback, Immersive Media, Virtual Reality

**ACM Reference Format:**
Vanessa Pfeiffer, Sebastian von Mammen, and Daniel Pohl. 2024. Haptic and Auditory Feedback on Immersive Media in Virtual Reality. In *30th ACM Symposium on Virtual Reality Software and Technology (VRST '24), October 9–11, 2024, Trier, Germany*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3641825.3689511

## 1 INTRODUCTION

As VR headsets become more popular and affordable, users can easily explore virtual environments for viewing images and videos. We explore enhancing the VR image viewing experience by integrating haptic and auditory feedback to better understand technical challenges and their impact on user engagement and immersion. Users explore 3D representations of images while experiencing vibrotactile feedback using controllers synchronized with audio feedback. The image is separated into different materials and the appropriate feedback is chosen based on the controller position of the user. We research the impact of synchronized haptic feedback and audio cues on the user's overall experience and immersion when interacting with VR media. The 3D representations of the images are created using depth maps and material maps. We address the challenges of utilizing depth maps from different image formats, including short-angled, 180°, and 360° images, in both 2D and 3D formats. Furthermore, we present a solution for mitigating weak spots in the depth maps of 360° images.

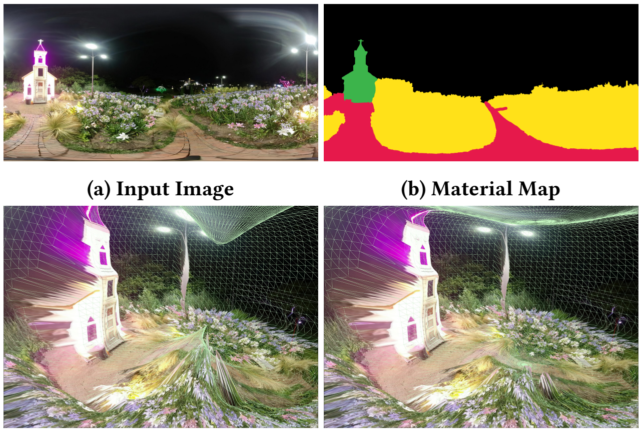
## 2 RELATED WORK

Estimating monocular depth using machine learning on planar images is a well-researched topic [7]. Large state-of-the-art deep learning models create accurate depth maps for a large variety of test sets. A problem in estimating monocular depth on spherical images is the distortion of the image that is often stored in an equirectangular format. Rey-Area et al. [9] counter this by projecting the image into multiple, less distorted images. These are converted to depth maps through deep learning models, and then patched back into one equirectangular image.

To enhance the immersion of a user when interacting with virtual materials, different feedback methods can be incorporated. Vibrotactile and visual cues can convey the sensation of compressing objects [10]. Devices with vibration motors to stimulate haptic feedback are available in different forms, most commonly hand controllers and haptic gloves [8]. While the addition of audio cues


*VRST '24, October 9–11, 2024, Trier, Germany*

ACM ISBN 979-8-4007-0535-9/24/10
https://doi.org/10.1145/3641825.3689511



(a) Input Image (b) Material Map

(c) Initial Mesh (d) Rectified Mesh

**Figure 2: (a) Equirectangular 360° image. (b) Designed material map. (c) Generated mesh from depth map with issues at the poles. (d) Fixed mesh at the poles with our algorithm.**

does not contribute to the haptic illusions, it adds to a perception of the naturalness and thus increases the measure of presence [6].

Touchly [1] is a closed-source app on the Meta store that uses depth maps to create touch feedback for VR180-3D media. However, they provide no information about their process and have not published any research on the effect on the user's experience.

## 3 METHODOLOGY

We extended the immersive media viewer immerGallery [3] to include depth and material maps and configuration data, using Zoe Depth [5] or Immersity AI [4] to generate these maps from our images. After preliminary testing for accuracy and speed, especially with 180° and 360° images, these tools were selected. Both models are state-of-the-art, delivering high-quality depth maps for planar and panoramic images quickly.

To enable touch interactions, colliders are generated from depth maps. For 180° and 360° media, vertices are placed in a spherical shape based on the depth given by the map. Planar images extrude the vertices of a plane that is curved based on the field of view of the image, along their normals based on the depth. The level of detail in these colliders is balanced against performance considerations to allow execution on standalone VR platforms such as Meta Quest 3. For most spherical images, this results in an elongated cylinder with foreground elements protruding inwards. A major challenge with spherical images in equirectangular projection is the significant distortion at the top and bottom. All evaluated depth algorithms inaccurately underestimate values, resulting in mesh spikes as shown in Figure 2c. Our solution recalculates depth and vertex positions in problem areas by estimating correct heights from surrounding vertices. Figure 2d shows the mesh after our correction.

The material maps define the areas of different haptic materials and are manually created using image editing tools. Each haptic material uses a unique color. Figure 2b shows a map for the example input image of Figure 2a. The haptic feedback of a material is provided by a sequence of frequency and amplitude, generated using Meta's Haptics Studio [2] to match to their correlating audio cues. During playback, the audio cues are synchronized with the vibration creating an integrated immersive experience. The sounds are chosen to closely resemble the interaction of a human hand with each material. When the user moves the controllers close or along the image collider, the material map is sampled at the point of interaction. This returns the haptic material for the related area and the correct audio cue and haptic sequence are played back. For example, if the user touches a cat, the material map would return the haptic material for it. This would lead to a purring sound with accompanying vibration being played.

## 4 RESULTS

The creation of depth maps for planar images works well. Many available depth estimation models deliver maps accurate enough to successfully enhance the image view experience. For panoramic, and especially 360° images, current models struggle with consistent, correct depth at the bottom and top. Our solution manages to successfully work around those inaccuracies and produce a usable 3D mesh. In our sample 360° images, the depth is adjusted by an average of 11 %, with 8 % of all vertices moved by more than 55 %. The method is effective unless the image has vertical protrusions at the poles.

In a short subjective evaluation, the feeling of presence was increased due to the haptic and auditory feedback.

## 5 CONCLUSION AND FUTURE WORK

In conclusion, this research suggests that the immersion of media viewing in VR could be enhanced with haptic and auditory feedback. Using depth maps and haptic materials is shown to be a reliable way to enable synchronized audio and haptic cues in the experience. Future work will focus on enabling interaction with distant objects through a virtual hand model and incorporating visual shaders to enable pseudo-haptic feedback methods. A user study will evaluate the impact of these features on the overall VR experience. We also plan to apply our research to videos, where frame-to-frame consistency is crucial.

## REFERENCES


[1] 2023. *Touchly*. https://touchly.app
[2] 2024. *Haptic Sutdio*. https://developer.oculus.com/resources/haptics-studio/
[3] 2024. *immerGallery*. https://immervr.com
[4] 2024. *Immersity AI*. https://www.immersity.ai/
[5] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. 2023. ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth. arXiv:2302.12288
[6] Frank Biocca, Yasuhiro Inoue, Andy Lee, and Heather Polinsky. 2002. Visual cues and virtual touch: Role of visual stimuli and intersensory integration in cross-modal haptic illusions and the sense of presence. (01 2002).
[7] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. 2021. Deep learning for monocular depth estimation: A review. *Neurocomputing* 438 (2021), 14–33. https://doi.org/10.1016/j.neucom.2020.12.089
[8] Tomas Novacek and Marcel Jirina. 2020. Overview of Controllers of User Interface for Virtual Reality. *PRESENCE: Virtual and Augmented Reality* 29 (12 2020), 37–90. https://doi.org/10.1162/pres_a_00356
[9] Manuel Rey-Area, Mingze Yuan, and Christian Richardt. 2022. 360MonoDepth: High-Resolution 360° Monocular Depth Estimation. In *CVPR*.
[10] Antony Tang, Mark Billinghurst, Samuel Rosset, and Iain Anderson. 2023. Enhancing Virtual Material Perception with Vibrotactile and Visual Cues. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 1011–1012. https://doi.org/10.1109/VRW58643.2023.00351