Immersive and Efficient 3D Gaussian Splatting in VR

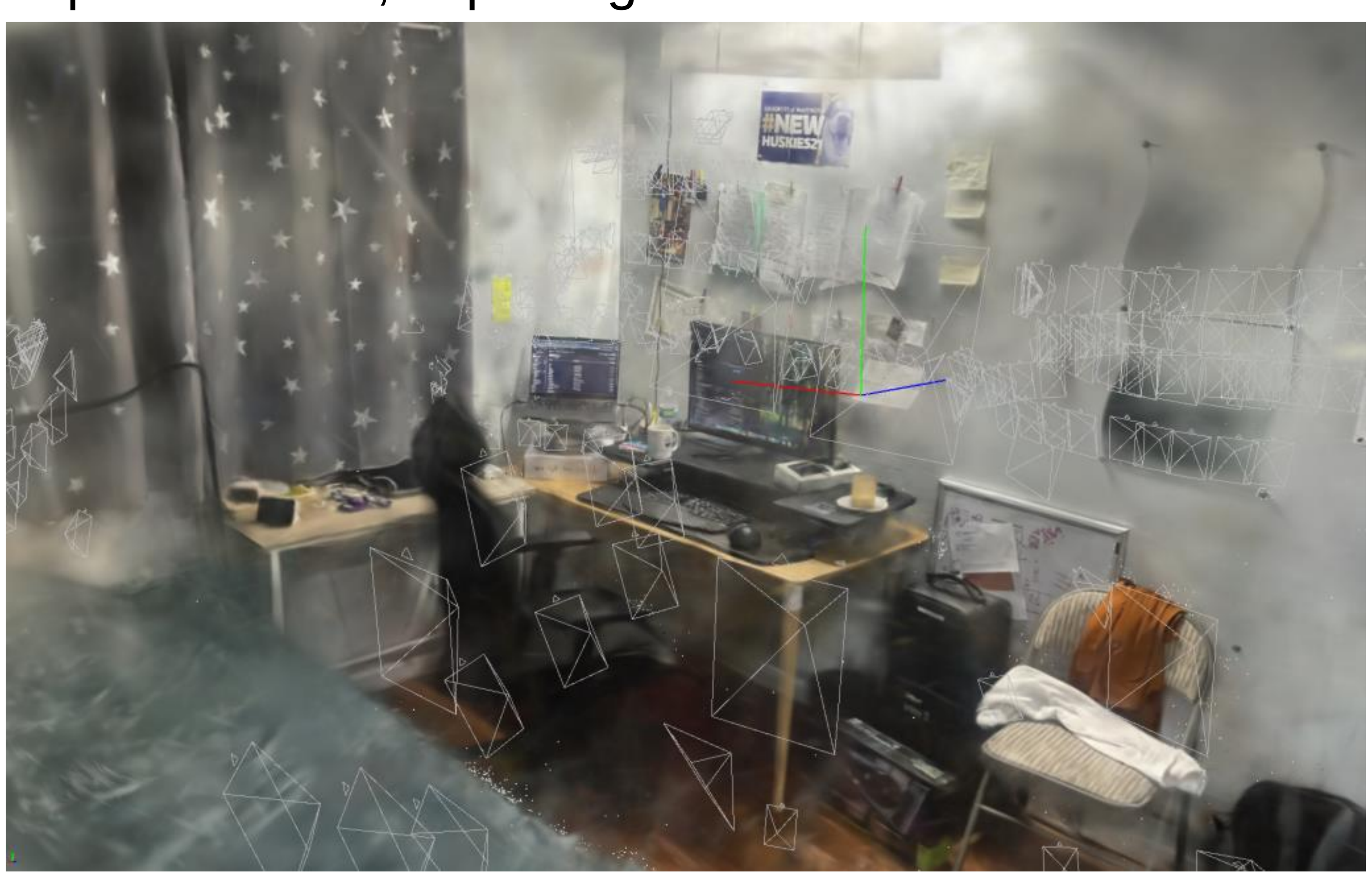
Derek Zhu, Hoang Nguyen, Rich Chen

PROJECT STATEMENT

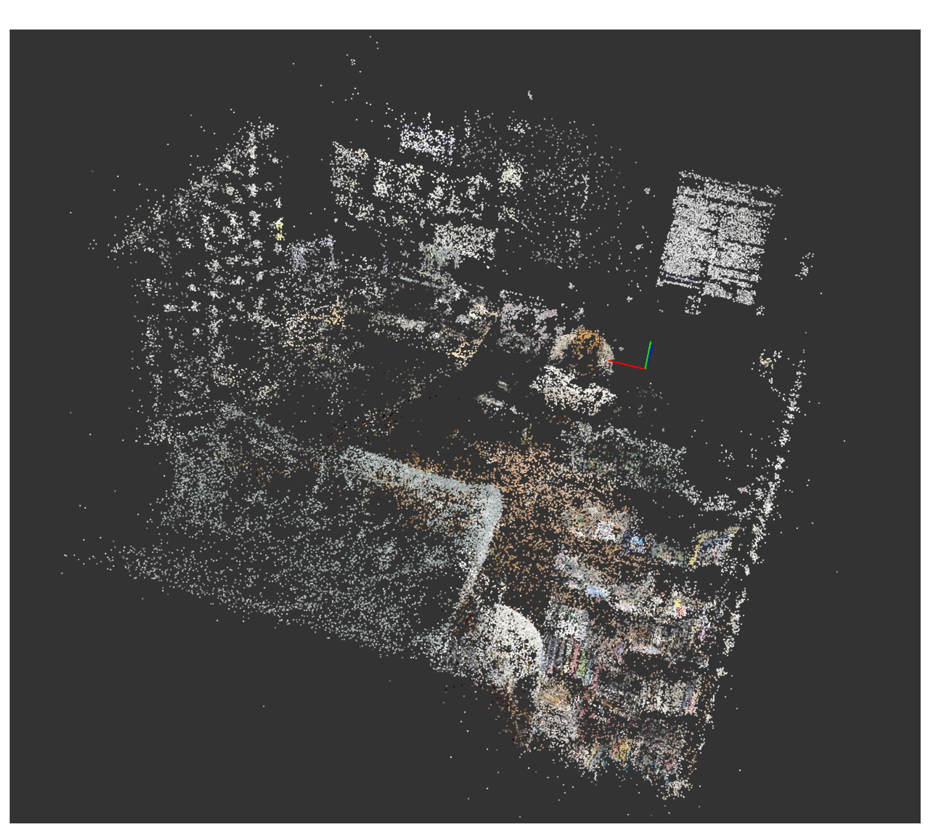
This project aims to create a wearable device and software platform that captures and recreates immersive, explorable VR memories using 3D Gaussian Splatting and spatial audio.

GAUSSIAN SPLATTING

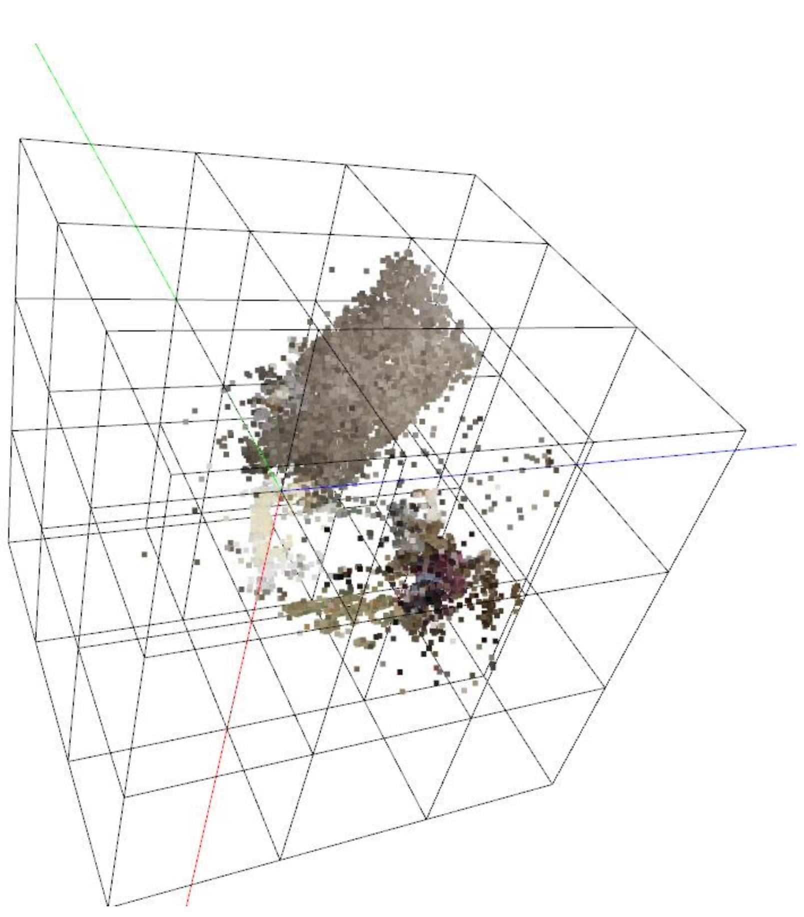
- 3D rendering technique that represents scenes using a collection of 3D Gaussian "splats" rather than traditional meshes or point clouds.
- Blend splats to get realistic-looking visuals.
- Allows use to combine many 2D images into a 3D representation, capturing the real world.



SPLATTING PIPELINE



- Major part of the splatting process is the get these point clouds.
- Our hardware allows us to get IMU readings to get relative positioning.
- Point clouds allow us to perform ray casting to identify how to merge Gaussian splats together.
- With a merging process defined, we can construct a database to store the merged Gaussian splats together, allowing users to query for particular positions.
- Splatting for time t+1 from time t and additional data:
  (allows us to add objects to a splat)

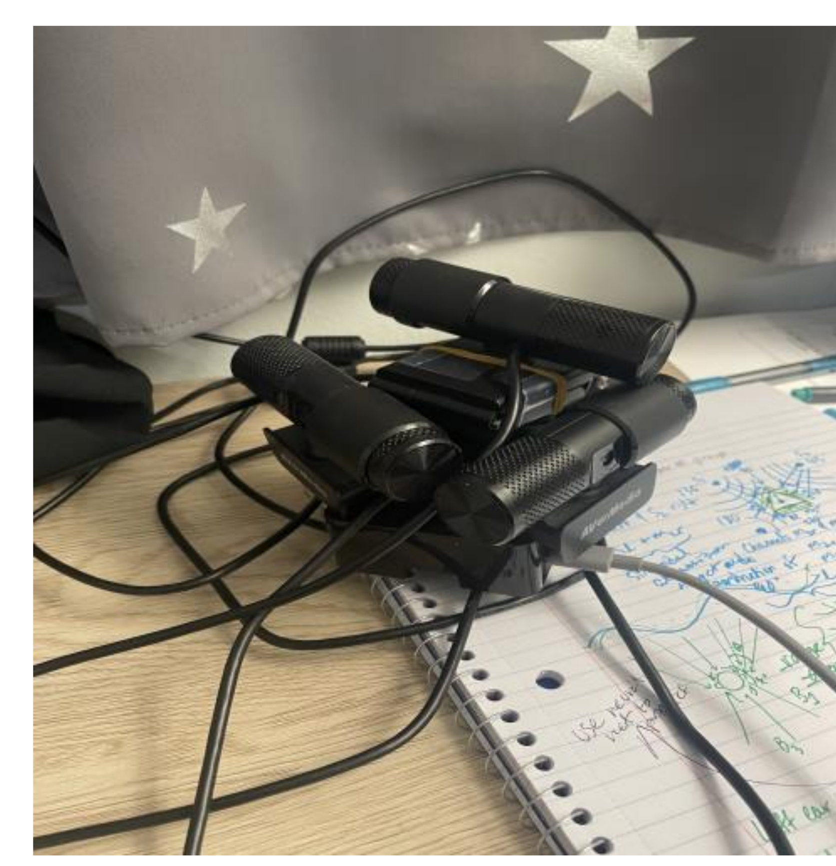


MAJOR CONTRIBUTIONS

1. A method for systematically chunking a capture into "blocks" and programatically determining all camera views associated with that block in order to cache and piece by piece later.
2. Method of splatting a new scene (t+1) where t is the time of the environment, utilizing the point cloud of scene (t) and images from (t+1).
3. A proof-of-concept for a wearable piece of hardware that captures images and audio of the environment of the user wearing it.
4. A method to produce spatial audio captured from the environment.

HARDWARE PROTOTYPE

- A constructed wearable device with 3 cameras and 6 microphones
- Could potentially capture 1 photo per second in three different directions, as well as 6 audio channels.
- Hardware for this is difficult to engineer – this is a major area for future work!



SPATIAL AUDIO

- Use direction of arrival (DOA) estimation.
- After microphones synchronized, must confirm time differences.
- Generalized Cross-Correlation with Phase Transforms (GCC-PHAT):

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} X_i(f) X_j^*(f) e^{j2\pi f\tau} \, df$$

$R_{ij}(\tau)$: Cross-correlation function between signals X_i and X_j ebign left and right respectively.
$X_i(f)$: Frequency-domain representation of signal i.
$X_j^*(f)$: Complex conjugate of the frequency-domain representation of signal j.
$e^{j2\pi f\tau}$: Complex exponential term that introduces a time shift τ.
τ : Time shift parameter.
f : Frequency variable.
df : Differential element in the integration over all frequencies.