Sensor2Scene: Foundation Model-driven Interactive Realities

Yunqi Guo[1], Kaiyuan Hou[12], Zhenyu Yan[1], Hongkai Chen[1],
Guoliang Xing[1], and Xiaofan Jiang[2]

[1]The Chinese University of Hong Kong, [2]Columbia University

Outline

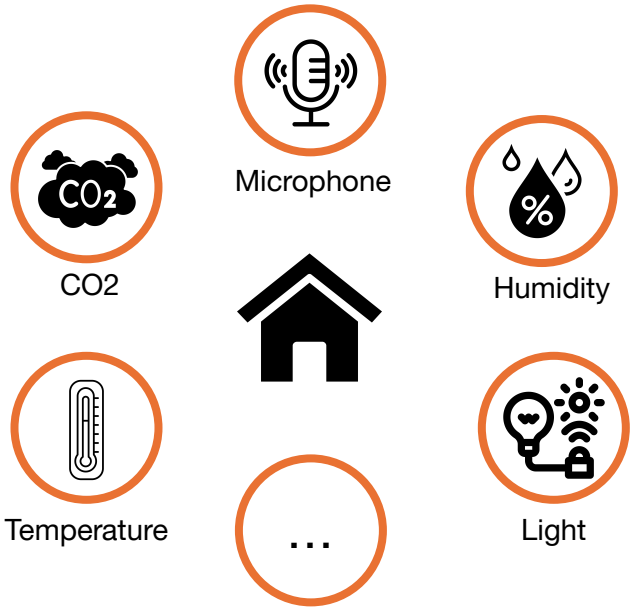
- Background & motivation
- Problem to solve: can we xxxxx?
- Existing solution and challenges
- Our solution design
  - Part1: connecting the sensor data with scene descriptions
  - Part2: create the visualization of the scene with text-to-3D
- Results: demonstrations, compare different text-to-3D models
- Discussions
  - Findings: our approach shows the potential
  - Limitations: current text-to-3D is not good at the abstract content represent, cannot handle dynamic content, etc
- After this paper and future work

https://www.notion.so/yqq/Presentation-b319950476f74c5184497b193fa5286a?pvs=4

Background: Sensor and Sensor Information

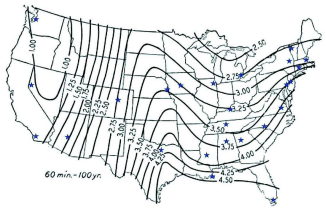
✓ Rich information on measuring the world

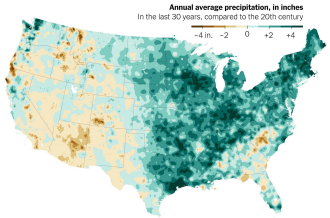
✗ Difficult for the average user to understand



Humidity:
58.4%
Light: 488.6 Lux
CO2: 633

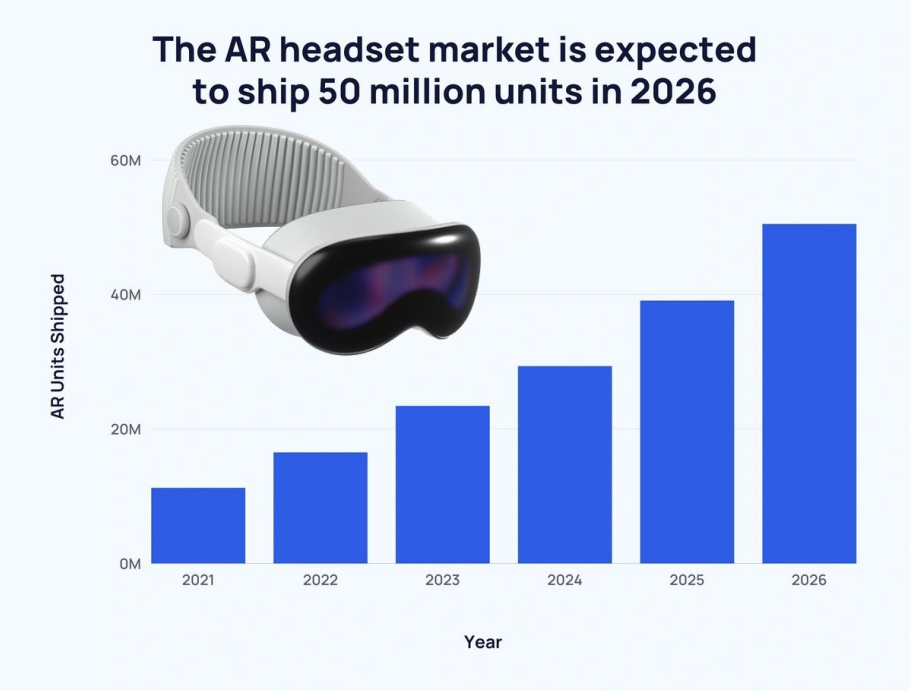
Numbers → Visualization





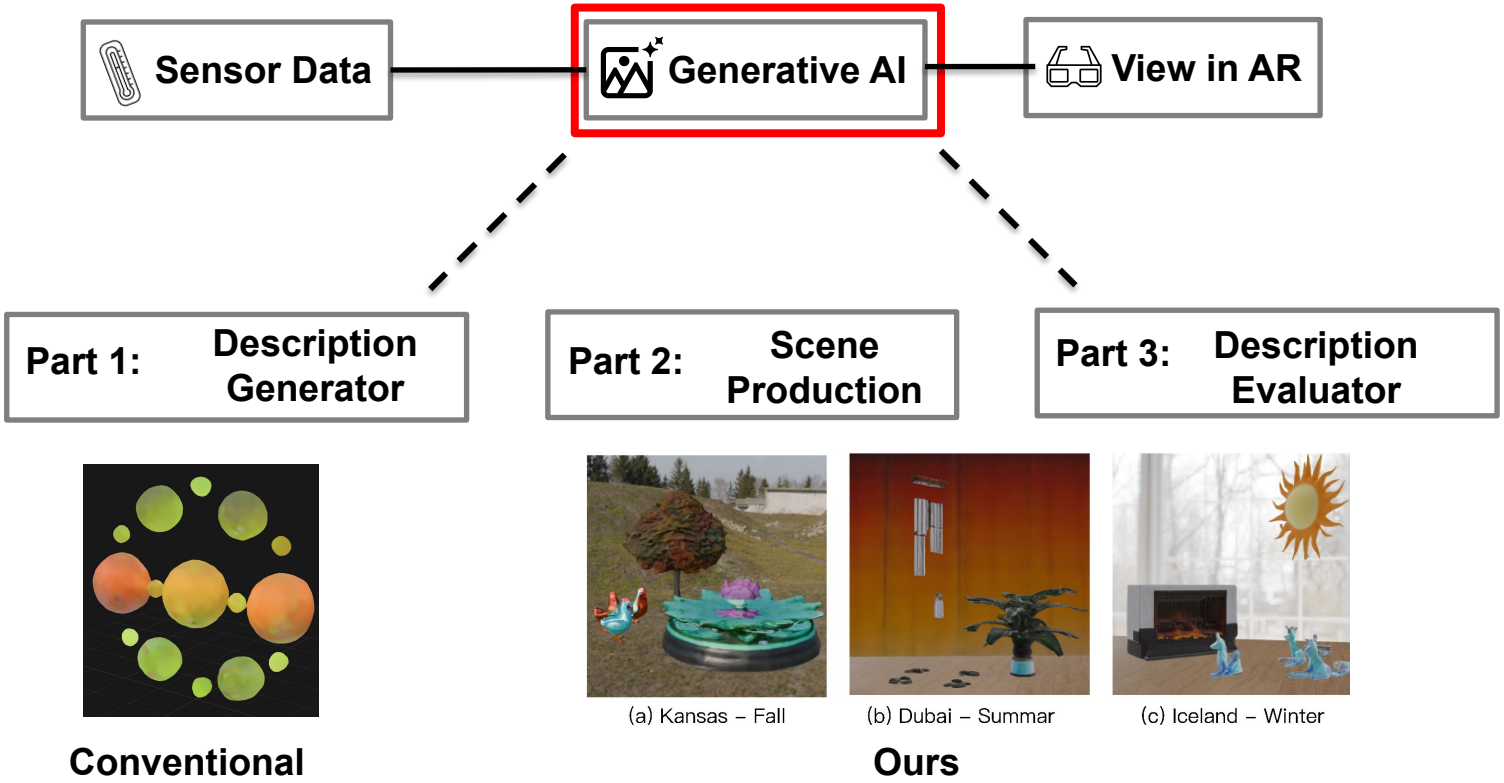
Precipitation Intensity

Expanding World of Augmented Reality (AR)



AR for Enhancing Human Perception and Interaction in the Reality

Sensor2Scene: Scene Generation from Sensor Data with FMs



Part 1: Scene Description Generator

Goal: sensor → scene description

Insights: LLM can be a powerful tool

- Zero or few shot classifier on sensory information
- Good for imagination and creations.

1. Input data → Interpretation

   Desired: number → perception
   Example: 39C → hot

Input

Temperature: {value: **39**,Unit: Celsius},
Wind: {value: **1.65**, Unit: m/s},
Location: Dubai, Time: 10:34 AM

Prompt

… Expert in interpretation of sensor data to determine how people might feel in those conditions. … {Input data}

Output

Temperature: High Heat: A temperature of 39°C is **quite high**… **Discomfort**…
Wind: …is considered a **light breeze** on the Beaufort wind scale.

Part 1: Scene Description Generator

Goal: sensor → scene description

Insights: LLM can be a powerful tool

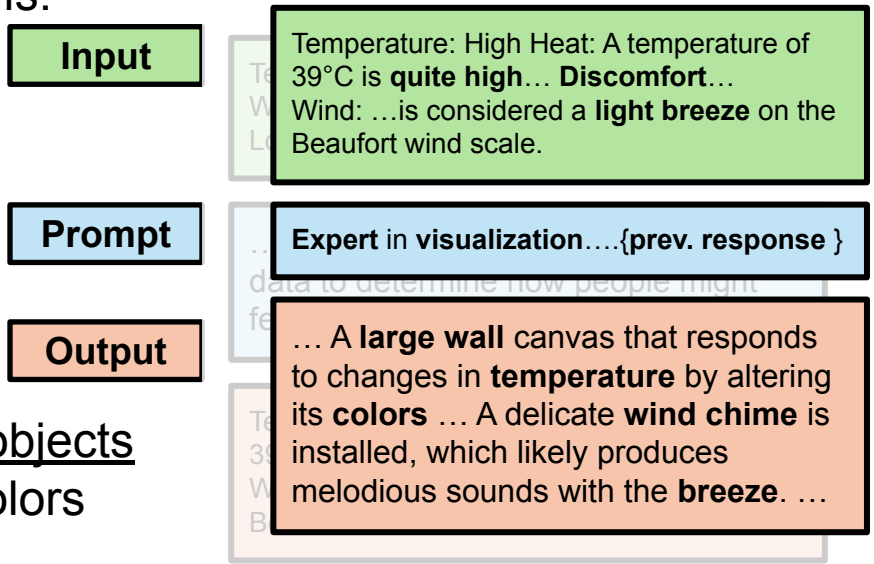
- Zero or few shot classifier on sensory information
- Good for imagination and creations.

1. Input data → Interpretation

   Desired: number → perception
   Example: 39C → hot

2. Interpretation → Scene Depicting

   Desired: perception → Tangible objects
   Example: hot → large wall with colors



Part 2: Text Description to 3D Scene

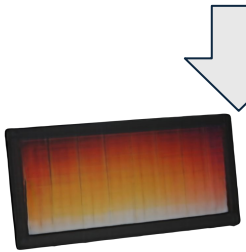
Problem: Convert Scene Descriptions to 3D Objects

Approaches

- **Approach 1:** Custom Programming ❓
  - High complexity; needs more programming skills.
- **Approach 2: Text to 3D Conversion** ✅
  - **Challenge:** Models support only individual objects.
  - **Steps:**
    i. Identify tangible objects and features (color, texture, size).
    ii. Create individual 3D objects.
    iii. Arrange objects in the environment.

> … A **large wall** canvas that responds to changes in **temperature** by altering its **colors** … A delicate **wind chime** is installed, which likely produces melodious sounds with the **breeze**. …

Scene Description from Part 1



A large red wall

A delicate wind chime

Part 3: LLM Evaluator in the loop

Observations:

1. Response can be too descriptive.
2. Ignoring part of input data

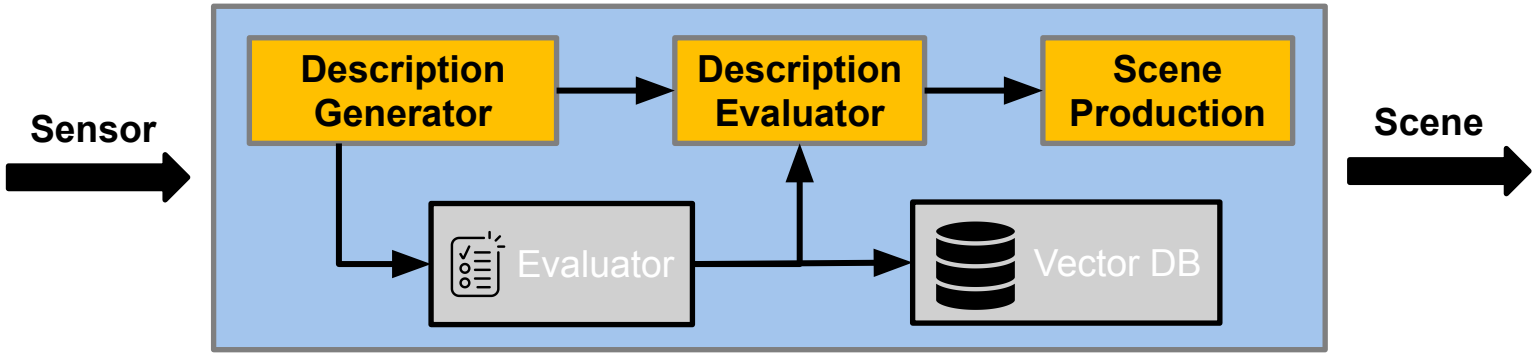
Evaluator: Scoring the scene description based on

- *Specificity, Utilization, Fidelity, Integration, Coherence*

Vector DB: Recording good descriptions for future reference

Input: Location: gym
Humidity: 55.6%; Temperature: 27.65

The gym environment could be visualized with **varying shades of red and orange to indicate the warmth of the space**... They would notice the colors shifting **subtly** to reflect the temperature changes within the gym…



System Implementation

Foundation Models

- Language Model: GPT-4
- Text-to-3D Model: Luma.AI[1] , DreamGaussian[2]

AR Interface

- Backend: Flask-AFrame Server
- Rendering: WebXR

Device Support

- Supported Devices: Meta Quest 3, Vision Pro

Device Setup: AR Glasses and Smart-home Sensor

[1] LumaAI: http://lumalab.ai
[2] Tang, Jiaxiang, et al. "Dreamgaussian: Generative gaussian splatting for efficient 3d content creation." arXiv preprint arXiv:2309.16653 (2023).

Synthesized Scene

(a) Kansas – Fall

Temperature: 22C
Humidity: 50%
Air Quality: 30AQI

(b) Dubai – Summar

Temperature: 40C
Humidity: 50%

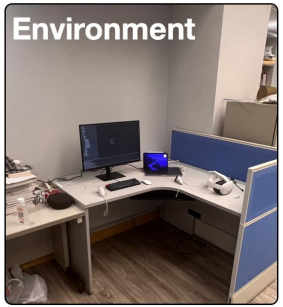
(c) Iceland – Winter

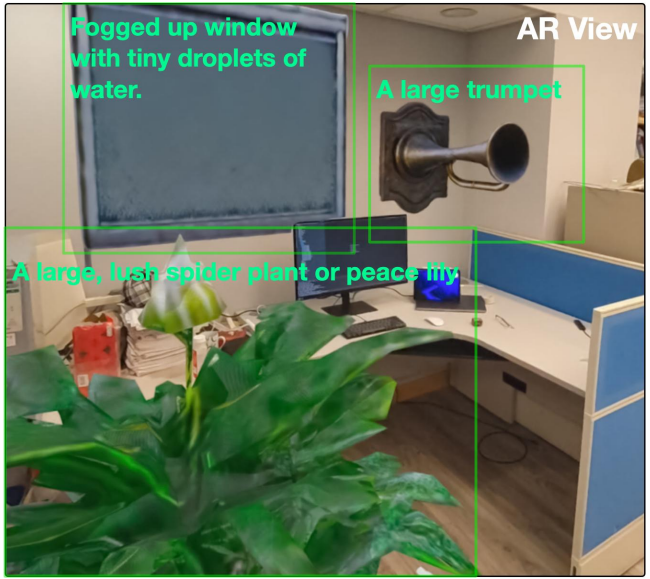
Temperature:8C
Light: 50Lux
Humidity: 40%

Result on Real World Scene

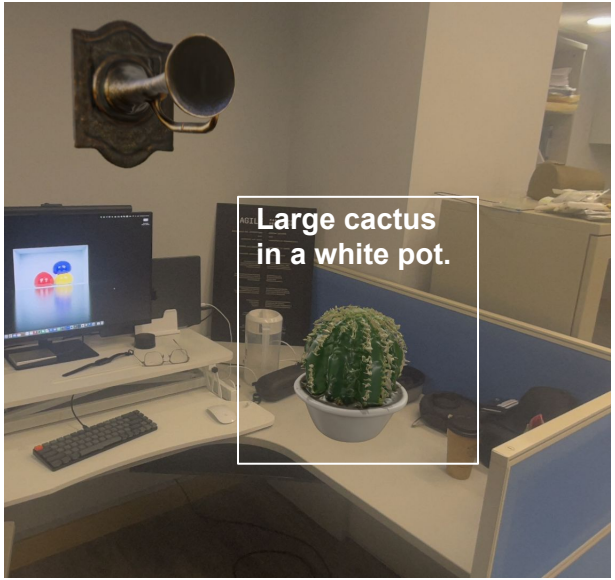
Sensor Input:

Temperature: 23 C
Humidity: 86%
Noise: 75 dB
Air Quality: 25 AQI





Example 1: humidity 86%, noise: 75 dB



Example 2: humidity 32%, noise 75 dB

Key Takeaways & Limitations

Takeaways

- **LLM/LMM:** Bridges sensor data with scene creation.
- **LLM Evaluation in the Loop:** Five metrics for assessing scene descriptions.
- **Pipeline:** Data capture to AR visualization.

Limitations

- **Text-to-3D:** Challenges with abstract sensor information like smoke.
- **Sensor Data Interpretation:** Risk of misinterpretation by LLMs.

Next Step 1: 3DGS for Smoke/Hot Weather

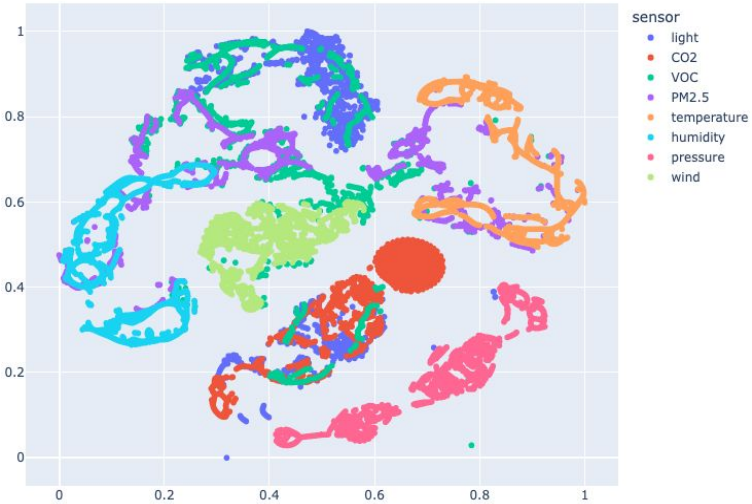
Objective: Can we draw the volume of hot air or smoke?

Solution:

- 3D Gaussian Splatting (3DGS)
- Learn the abstraction content features, e.g., volume, colors, dynamics, etc.)

Next Step 2: Training Sensor Data Embedding with Visual Modality

Original **Masked**



3DGS ⇔ Visualization Space ⇐ Sensor Data

Thank You / Questions?

Kaiyuan Hou

Columbia University

Email: kh3119@columbia.edu