Demo Abstract: PixelGen: Rethinking Embedded Camera Systems for Mixed-Reality

Kunjun Li*
National University of Singapore
kunjun_li@u.nus.edu

Manoj Gulati*
National University of Singapore
manojg@nus.edu.sg

Dhairya Shah
National University of Singapore
dhairya@nus.edu.sg

Steven Waskito
National University of Singapore
steven.waskito@u.nus.edu

Shantanu Chakrabarty
NCS Group
shantanu.chakrabarty@ncs.com.sg

Ambuj Varshney
National University of Singapore
ambujv@nus.edu.sg

Figure 1: PixelGen rethinks the architecture of embedded camera systems. PixelGen captures a broad representation of the world through an array of sensors and transceivers. The simplicity of the sensor and low-resolution imagery, facilitates low-power operation. The captured data is then processed using a transformer-based diffusion model, allowing the generation of novel environmental representations. This includes the generation of high-definition images.

ABSTRACT

A confluence of advances in several fields has led to the emergence of mixed-reality headsets. They can enable us to interact with and visualize our environments in novel ways. Nonetheless, mixed-reality headsets are constrained today as their camera systems only capture a narrow part of the visible spectrum. Our environment contains rich information that cameras do not capture. It includes phenomena captured through sensors, electromagnetic fields beyond visible light, acoustic emissions, and magnetic fields. We demonstrate our ongoing work, PixelGen, to redesign cameras for low power consumption and to be able to visualize our environments in a novel manner, making some of the invisible phenomena visible. PixelGen combines low-bandwidth sensors with a monochrome camera to capture a rich representation of the world. This design choice ensures information is communicated energy-efficiently. This information is then combined with diffusion-based image models to generate unique representations of the environment, visualizing the otherwise invisible fields. We demonstrate that together with a mixed reality headset, it enables us to observe the world uniquely.

1 INTRODUCTION

Rapid advances in several areas are leading to the design of new computing platforms that are now motivating us to rethink the fundamental task of the camera. One of these emerging computing platforms is mixed reality headsets like Apple Vision Pro and Meta Quest 3. These headsets are full-fledged computers with various sensors, computing, and communication capabilities that process information and visualize the environment in front of our eyes on a high-resolution display. This platform allows us to interact with our environment and visualize and understand it in new ways.

*These authors contributed equally to this work and are co-primary authors.

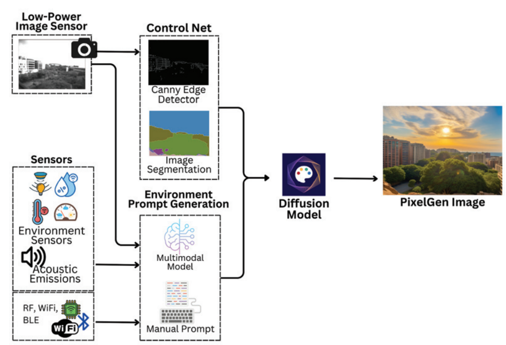

Figure 2: PixelGen communicates rich sensor information energy-efficiently to an edge device. Edge device uses diffusion models to generate a rich representation of the environment. This enables us to visualize information that conventional cameras do not capture.

Mixed reality headsets open up several application scenarios. Our physical world has sensors capturing vast amounts of information, including environmental conditions and radio waves. Projecting this information onto mixed reality headsets and overlaying it on objects could help us visualize it in novel ways. For example, a sensor in a plant could transmit information to a headset, giving real-time data about plant health. We may also generate creative representations of the environment for virtual sessions.

Nonetheless, it is challenging to support scenarios mentioned above. The reason is that the camera systems employed in mixed-reality headsets only capture a narrow representation of the world illuminated by visible light. Furthermore, cameras are energy consuming, which makes it difficult to continuously stream information wirelessly from deployed cameras in the environment to the headset unless cameras are tethered to an external power supply.

979-8-3503-6201-5/24/$31.00 ©2024 IEEE

DOI 10.1109/IPSN61024.2024.00036

(a) Monochrome Image (b) Segmented Output (c) Oil Painting (d) Realistic Image (e) User with Headset

Figure 3: (a) Monochrome image (324x244) captured by PixelGen, (b) Location and shape of objects extracted through ControlNet, (c) Representation of environment as an oil painting, (d) Representation of a realistic and high-resolution image, (e) A user visualizing environment representation generated by PixelGen on a mixed-reality headset, Xreal Air 2.

Cameras have evolved over thousands of years, from the camera obscura that projected light onto a surface to aid artists, to contemporary digital cameras [1]. Despite these advancements, cameras still capture only a narrow part of the world illuminated by visible light. We demonstrate our ongoing work on PixelGen, which rethinks camera architecture. PixelGen captures broad environmental parameters using a low-resolution image sensor and low-bandwidth sensors. It then transmits the captured information energy-efficiently [2, 3], enabling low power consumption for the camera platform. This can enable the PixelGen camera to even operate for prolong periods on energy derived from a small battery.

Furthermore, despite capturing low-resolution and bandwidth information, PixelGen generates rich representations of the environment, as shown in Figure 1 & 2. It uses state-of-the-art diffusion models to combine various sensor information and generate high-definition images. These images visualize phenomena hidden in conventional cameras on the mixed-reality headsets.

2 DESIGN

PixelGen features a custom hardware platform, a mixed-reality headset, and an edge computer. The platform captures broad sensor information, including low-resolution images. It communicates this sensor information energy-efficiently using a tunnel diode or backscatter transmitter [3]. The system uses the sensor information with a language model to transform sensor data into prompts, which are then used with a diffusion model to generate visual representations of the environment. Figure 2 shows an overview of the system. We have designed the platform on an FR4 substrate, a 4-layer PCB manufactured by OSHPark. We show this in Figure 1.

Platform. We design the PixelGen hardware to enable prolonged operation on a small battery or energy harvested from the environment. The platform's core is the highly energy-efficient Ambiq Apollo3 Blue microcontroller. We equip the platform with a Bosch BME280 to measure temperature, humidity, and pressure information from the environment. We estimate the ambient light conditions using a ROHM BH1750 sensor. The platform also has a low-resolution monochrome image sensor, Himax HM01B0.

Edge device. The system then processes the collected sensor data using language and diffusion-based image models. The language model interprets the collected sensor data and generates appropriate prompts, which become the input for the diffusion model. We can then visualize the generated images from the diffusion model on the mixed-reality headsets. Even though the collected information consists of monochrome images and low-bandwidth sensor data, the diffusion model can combine them to generate a high-resolution representation of the environment. We leverage Stable Diffusion and OpenAI ChatGPT for the image and language models running on a dedicated edge device. Nonetheless, as these models become increasingly capable, they can run on constrained devices. We expect these models to be able to run on mixed-reality headsets, enabling cameras to communicate directly with the headset.

Implementation and early results. We conduct various experiments to evaluate our system. We capture an image using the platform's monochrome camera and the remaining environmental data. We then process these and use them with the diffusion model to generate rich representations of the environment. Figure 3 shows various images generated by the model and a user visualizing them. These representations include higher-resolution representations of the environment and more unique and creative representations, such as environments visualized as oil paintings.

3 DEMONSTRATION

We plan to demonstrate PixelGen together with a mixed reality headset from XReal. In this demonstration, PixelGen will capture images and other sensor data from the environment. Next, we will transform these sensor data into prompts. We will demonstrate the various sensor data, images, and prompts generated by our system. Then, we will leverage the diffusion model to generate high-resolution and creative images based on the sensor data and prompts. We will visualize and project these images on the extended reality headset. Users will be able to visualize their surroundings in various ways in real time, demonstrating our system. During the demonstration, we will also explain to attendees the basics of our system and the advantages of our system from the perspective of bandwidth and power consumption.

ACKNOWLEDGMENTS

This work is partially funded through a startup grant (A-8000277-00-00), a grant from the NUS-NCS centre (A-0008542-02-00), and MoE Tier 1 Grant (A-8001661-00-00) hosted at the National University of Singapore.

REFERENCES

[1] G. A. Lloyd. 1977. Electronic still Camera. https://patents.google.com/patent/US4131919A/en. [Online; accessed 13-Oct-2023].

[2] A. Varshney et. al. 2017. LoRea: A Backscatter Architecture That Achieves a Long Communication Range *(ACM SenSys'17)*.

[3] A. Varshney et. al. 2022. Judo: addressing the energy asymmetry of wireless embedded systems through tunnel diode based wireless transmitters. In *ACM MobiSys'22*.