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

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

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

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

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

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

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

ABSTRACT

The confluence of advances in several fields is leading to the emergence of new computing systems, such as mixed-reality headsets. These systems with rich sensors, computing, and networking capabilities project information on a display very close to the eyes. They can enable us to interact with and visualize our environments in novel ways. Nonetheless, today, these headsets are constrained as the camera systems they interface with only capture a narrow part of the visible spectrum. However, our environment consists of a rich system of devices capturing information through sensors, electromagnetic, acoustic, and radio fields, as well as other phenomena. Furthermore, cameras are power-consuming, which means they have a limited lifespan on batteries and are more typically tethered to an external power supply. We demonstrate our ongoing work to redesign cameras for low power consumption and to be able to visualize our environments in a novel manner. PixelGen combines low-bandwidth sensors with a monochrome camera to capture a rich world representation. This ensures low-bandwidth capture of information, which is communicated in an energy-efficient manner through a low-power transmitter, thus ensuring prolonged operation even on a small battery. This information is then combined with diffusion-based image models to generate unique representations of the environment. We demonstrate that the image representations could enable the user to uniquely visualize the environment.

1 INTRODUCTION

Rapid advances in several areas, such as display electronics, communication, and computation, 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 situated very close to the eyes. This platform allows new ways for us to interact with our environment and visualize and understand it.

The enhanced capabilities offered through mixed reality headsets open up several application scenarios. Our physical world is distributed with sensors capturing vast amounts of information, including environmental conditions and radio waves. Projecting such information onto mixed reality headsets and overlaying them on objects could help us visualize them in a novel way. For example, a sensor embedded in a plant could transmit information to a mixed-reality headset, which could be overlaid to give us real-time information about plant health. There are many other scenarios possible. We may be able to generate novel and creative representations of the environment for our sessions in virtual spaces.

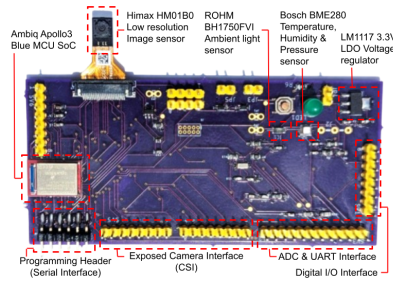

Figure 1: PixelGen platform captures a broad representation of the world through sensors and monochrome camera. They are communicated in an energy-efficient manner to an edge device or mixed reality headset. This is then used to generate unique representations of the environment.

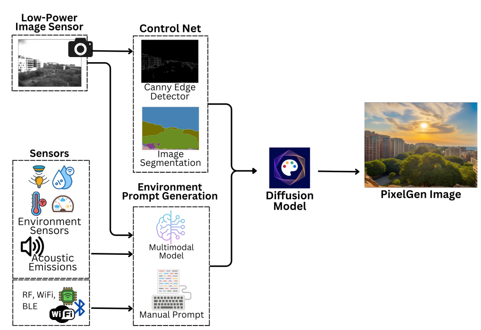

Figure 2: PixelGen system collects and processes the sensor data in a series of steps leading to generation of high-resolution and unique representations of the environment.

Nonetheless, it is challenging to support scenarios such as those 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 also energy-intensive, making it difficult to continuously stream information wirelessly from deployed cameras in the environment.

(a) (b) (c) (d) (e)

Figure 3: (a) Monochrome image (324x244) captured by the camera, (b) represents location and shape of various objects extracted through ControlNet, (c) represents a representation of environment as an oil painting generated by our system, (d) and (e) represent realistic and high-resolution image generated by the system.

Cameras are one of the most widely deployed embedded systems. They have evolved over thousands of years, undergoing developments beginning with the camera obscura [1], which projected light onto a surface to aid artists, and evolving to the Daguerreotype camera in the early 1800s [3], which captured images on silver-coated copper plates. The modern foundation for contemporary digital cameras was laid by Steven Sasson in 1975 at Kodak [2]. Despite these advancements, the inherent nature of cameras to capture the world illuminated by visible light has largely remained unchanged.

In this demonstration, we present our ongoing work to design PixelGen, which is a retake on the architecture of cameras. We design a novel and low-power camera that captures many parameters from the environment using low-resolution, low-bandwidth sensors. This information is communicated in an energy-efficient manner using tunnel diode-based transmitters [4]. The captured data is then used with state-of-the-art diffusion models to generate various representations of the environment, which are then visualized through the mixed-reality headset.

2 DESIGN AND EARLY RESULTS

PixelGen features a custom hardware platform called PixelSense , a mixed reality headset, and possibly an edge computer. The PixelSense platform captures various environmental data, including low-resolution images, temperature, humidity, light intensity, and radio, magnetic, and acoustic emissions. This data is transmitted to the edge device energy-efficiently using a Judo transmitter. The collected sensor information is then used 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.

Camera platform. The PixelSense hardware is designed 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. Wireless communication may be implemented using the Judo mechanism or a backscatter system like LoRea. We have designed the platforms on an FR4, 4-layer PCB manufactured by OSHPark. We show this in Figure 1.

Generating images and environment:. The collected sensor data is then processed using a 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. The generated images from the diffusion model can then be visualized on the mixed reality headsets. Even though the collected information consists of monochrome images and low-bandwidth sensor data, the diffusion model can put them together to generate a very 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 are becoming increasingly capable, they can also 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.

Putting everything together and early results. We conduct various experiments to evaluate our system. However, we present only one of these results and leave others for a detailed demonstration at the conference. We capture an image using the platform's monochrome camera and the remaining environmental data. These are then processed and used with the diffusion model to generate rich representations of the environment. Figure 3 shows various images generated by the model. These 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 particular 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 a real-time manner, 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.

REFERENCES

[1] Scott Billings. 2023. [Online; accessed 13-October-2023].

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

[3] Bates Lowry and Isabel Barrett Lowry. 2000. *The Silver Canvas: Daguerreotype Masterpieces from the J. Paul Getty Museum.* Getty Publications.

[4] Ambuj et. al. Varshney. [n. d.]. Judo: addressing the energy asymmetry of wireless embedded systems through tunnel diode based wireless transmitters. In *ACM MobiSys'22.*