Point, Detect, Discover

Interactive Annotations with Live Detection in XR

Arushi Jeloka, Kriti Sharma, Krishna Panchap

Problem

In dynamic professional, educational, and industrial environments, effective communication about physical spaces and objects is challenging —particularly when providing clear, contextual information for training, instruction, or collaborative tasks.

Our Approach

We leveraged the Meta Quest headset to enable intuitive, real-time annotations directly within physical spaces, providing clear, contextual information for training, instruction, and collaboration through immersive mixed-reality interactions.

Motivation / Use Cases

This has the potential for many diverse applications like in warehouses for inventory management, labs and medical centers for equipment and safety training, workshops like carpentry for tool instructions, and industrial and retail settings for maintenance and customer service to streamline operations and onboarding.

Our Process

Our AR annotation framework consists of three core components:

1. Real-time Object Detection: Utilizing YOLO, the system continuously processes captured screen frames to detect and track objects.
2. Gesture-based Selection: Hand-tracking via Meta's SDK enables intuitive gesture-based interactions, including gaze-based object selection and pointing gestures.
3. Contextual Annotation: Once an object is selected, cropped image data is passed to an on-device LLaVA model (34B) to generate annotations comprising object name, usage, and price range.

Caption: Annotations added by user using our AR application on the Meta Quest

Results

Current limitations include reliance on manual annotations, lack of automated object recognition, and the inability to provide video-based or fully immersive instructional content. Future development aims to integrate a vision-language AI model (VLLM) to enable real-time object detection and automatic annotations directly within the mixed reality environment.

Additionally, incorporating video annotations and immersive tutorials will greatly enhance the platform's value, especially for complex settings such as medical training facilities, mechanic workshops, or other technically demanding fields.

References

- https://developers.meta.com/horizon/develop/unity https://docs.unity3d.com/6000.0/Documentation/Manual/AROverview.html
- https://developers.meta.com/horizon/downloads/package/meta-xr-sdk-all-in-one-upm/
- https://huggingface.co/liuhaotian/llava-v1.6-34b
- https://huggingface.co/liuhaotian/llava-v1.6-34b