Demo: Liquid Identification via Vision-Guided mmWave Imaging and LLM Reasoning

Bo Liang[ℙ], Jingzhe Peng[𝔹], Xingyuming Liu[ℙ], Chen Gong[ℙ], Chenren Xu[ℙ𝕂✉*]
[ℙ]School of Computer Science, Peking University; [𝔹]Beijing Jiaotong University
[𝕂]Key Laboratory of High Confidence Software Technologies, Ministry of Education (PKU)

ABSTRACT

We introduce ErLang Sight, a novel multimodal system designed for liquid identification, integrating vision-based object detection, millimeter-wave (mmWave) Synthetic Aperture Radar (SAR) imaging, and large language model (LLM)-based contextual reasoning. Initially, the system leverages a visual detection pipeline to identify potential liquid containers within the environment, subsequently directing a mmWave sensor to perform targeted SAR imaging of these identified regions and the dielectric constants of the liquids are estimated using reflection coefficient analysis techniques. These physical measurements, combined with visual context and environmental indicators (such as whether the scenario is a kitchen, laboratory, or bar), are then input into a pre-trained LLM. The LLM employs advanced semantic and situational reasoning to accurately determine the most likely type of liquid by integrating physics-based data with contextual knowledge. Experimental evaluations demonstrate that ErLang Sight significantly enhances the accuracy of distinguishing visually ambiguous liquids and exhibits robust generalization to previously unseen environments.

1 INTRODUCTION

Liquid identification plays a critical role in numerous real-world applications, including water quality monitoring, spoilage detection in beverages, counterfeit liquid recognition, and hazardous substance identification. Traditional identification methods typically rely on cumbersome and costly equipment, restricting their practicality for everyday use. Recently, pervasive wireless technologies—such as Wi-Fi [1] and mmWave [2, 3]—have emerged, offering more accessible, compact, and cost-effective solutions for liquid sensing.

Despite these advancements, existing wireless-based methods still lack the ability to incorporate contextual understanding and reasoning. This limitation manifests in two significant ways: firstly, when multiple containers containing different liquids are present, current systems struggle to accurately match specific received signals to their corresponding containers, resulting in misidentification. Secondly, these approaches depend heavily on pre-collected datasets and rigid classification models, constraining their adaptability to new or changing environmental conditions and hindering their effectiveness in open-set recognition scenarios.

To overcome these limitations, we introduce ErLang Sight, a multimodal system for liquid identification that integrates mmWave imaging, vision-based perception, and large language model (LLM)-driven contextual reasoning. ErLang Sight harnesses visual detection to guide mmWave SAR imaging, which isolates the specific mmwave signals of the region of interests and ger rid of the interference and multipath. ErLang Sight also employs the semantic and situational reasoning capabilities of LLMs to achieve robust, accurate liquid classification by explicitly exploiting human instructions, environmental information provided in the image, characteristics of the liquid in the image, and dielectric constant estimated from mmWave, even in visually ambiguous scenarios and previously unseen environments. This innovative fusion significantly enhances the reliability and realism of liquid identification, paving the way for future developments in versatile, context-aware sensing systems and marking a substantial advancement toward intelligent material perception in real-world applications.

2 SYSTEM DESIGN

2.1 Vision-based Container Detection

The vision module in ErLang Sight utilizes an RGB-D camera, continuously active to monitor liquids within its field of view. It employs a YOLO-based detector specifically trained to recognize potential liquid containers such as bottles, cups, and bowls. Upon detection, each container is assigned a unique identifier, its location is tracked, and its bounding box is visually marked on the output image.

2.2 SAR-based mmWave Imaging

When a new container is identified by the vision module, the mmWave imaging module is activated. Our system uses a two-dimensional synthetic aperture radar (SAR), wherein the radar moves along both the x-axis and y-axis, coherently

*✉: chenren@pku.edu.cn

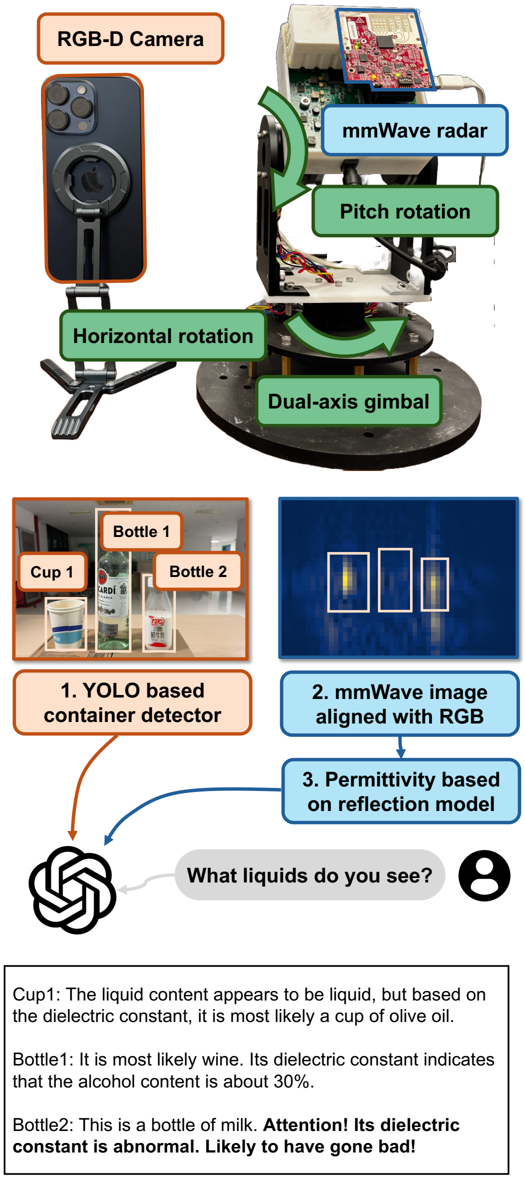

Figure 1: ErLang Sight Workflow.

integrating signals to form high-resolution images. To mitigate the inefficiency of scanning the entire spatial region, our design leverages the specular reflection characteristics of mmWave signals interacting with everyday containers. Specifically, we restrict radar movement to the container's specular reflection region—approximately within a 40-degree field-of-view in both azimuth and elevation angles. This targeted approach significantly reduces scanning time by more than threefold per container. The resulting radar images are then aligned precisely with the corresponding RGB images. Subsequently, the permittivity of the contained liquid is calculated using the dual-reflection model proposed in [3].

2.3 LLM-based Reasoning

One straightforward approach to infer liquid types involves comparing measured permittivity values against known values in reference tables. However, permittivity values are inherently variable, influenced significantly by composition and temperature. For instance, low-fat milk exhibits permittivity ranging from approximately $69.78 + 19.11j$ to $54.09 + 10.59j$, which overlaps considerably with permittivity values of strawberry juice and methanol-water mixtures (approximately $60 + 10j$), rendering direct differentiation challenging.

To address this ambiguity, our system integrates visual input and the common-sense reasoning capabilities of LLMs. For example, while methanol-water mixtures have toxic properties and are unlikely to be found in typical kitchen environments (as visually inferred), strawberry juice visually differs significantly from milk, providing additional contextual clues. Another scenario involves user-directed instructions to the LLM, such as "Determine if the milk has spoiled," effectively transforming the identification task into a simplified binary classification. The LLM's capacity to automatically combine visual context with common-sense reasoning presents an efficient, adaptable solution to liquid identification tasks.

3 DEMONSTRATION

The prototype and workflow of ErLang Sight are illustrated in Fig. 1. The demonstration setup involves placing various liquid containers, filled with different liquids, at distances ranging from 10 cm to 40 cm from the sensor on a desk environment. The demonstration encompasses several practical cases: (1) beverage type identification, where ErLang Sight distinguishes between visually similar drinks such as different types of juices or sodas; (2) milk spoilage detection, highlighting the system's capability to discern freshness based on the liquid's permittivity changes; and (3) alcohol content detection, emphasizing ErLang Sight's precision in identifying alcoholic beverages and estimating their alcohol concentration. These scenarios showcase the system's accuracy, reliability, and adaptability in realistic conditions, effectively illustrating its potential utility in everyday environments and specialized settings.

REFERENCES

[1] Fei Shang, Panlong Yang, Yubo Yan, and Xiang-Yang Li. Packquid: In-packet liquid identification using rf signals. *ACM IMWUT*, 2023.
[2] Yumeng Liang, Anfu Zhou, Huanhuan Zhang, Xinzhe Wen, and Huadong Ma. Fg-liquid: A contact-less fine-grained liquid identifier by pushing the limits of millimeter-wave sensing. *ACM IMWUT*, 2021.
[3] Zhu Wang, Yifan Guo, Zhihui Ren, Wenchao Song, Zhuo Sun, Chao Chen, Bin Guo, and Zhiwen Yu. Liqdetector: enabling container-independent liquid detection with mmwave signals based on a dual-reflection model. *ACM IMWUT*, 2024.