ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Toward Multimodal Liquid-Level Estimation for Closed-Loop Robotic Pouring

Hongyu Deng and He Chen
Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
Email: {hydeng, hechen}@ie.cuhk.edu.hk

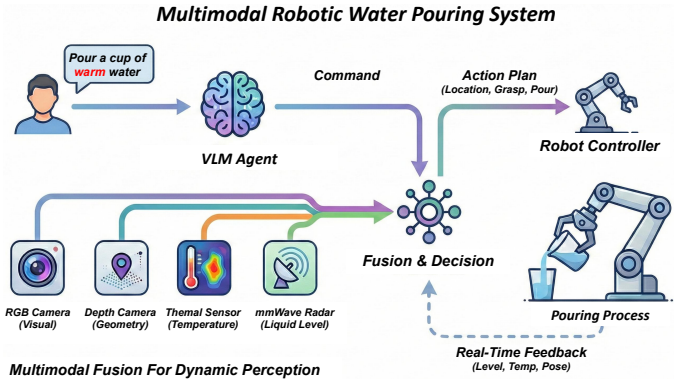


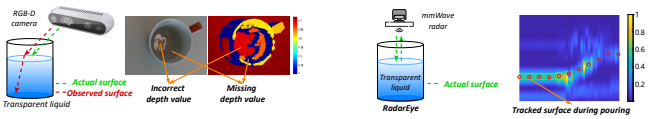
Fig. 1. Fast-slow architecture for robotic pouring. The VLM-based slow system handles task reasoning; the fast system provides real-time multi-sensor control.

***Abstract*—We consider the problem of real-time liquid-level estimation for closed-loop robotic pouring. To this end, we propose a fast-slow architecture where a Vision-Language Model handles high-level task reasoning and a sensor-driven fast system provides low-latency feedback. As a first instantiation of the fast system, we present RadarEye, a mmWave radar signal processing pipeline that tracks liquid level during pouring. RadarEye combines (i) AoA–ToF beamforming for liquid surface localization with (ii) a physics-informed tracker that suppresses multipath interference. In real-robot experiments, RadarEye achieves 0.35 cm median error at 0.62 ms per-update latency, outperforming vision and ultrasound baselines.**

I. Introduction & Motivation

Robotic pouring involves two distinct requirements: high-level task reasoning (e.g., recognizing containers, planning strategies) and low-latency reactive control during the pour itself. We propose a **fast-slow architecture** (Fig. 1) to separate these concerns. The *slow system* uses a Vision-Language Model (VLM) for semantic understanding and strategy planning. The *fast system* fuses multiple sensors (radar, RGB, thermal) to provide real-time feedback for closed-loop pouring control.

Within the fast system, a central problem is **reliable liquid level feedback**. RGB-D cameras produce missing or biased depth on transparent liquids due to specular reflection and refraction [1], [2] (Fig. 2(a)). Ultrasonic sensors transmit through liquid rather than reflecting off its surface. Vision-based neural methods [3] are sensitive to lighting and add ∼40 ms inference latency. Acoustic methods [4] infer level indirectly from flow sound, limiting accuracy.



(a) Depth errors on transparent liquid.

(b) RadarEye uses mmWave radar for level sensing.

Fig. 2. (a) RGB-D depth errors on transparent liquid. (b) RadarEye measures liquid level via radar surface reflections.

As a **first step** toward the multi-sensor fast system, we use mmWave radar for liquid level sensing. At 60 GHz ($\lambda \approx 5$ mm), transparent liquids reflect electromagnetic waves [2], enabling direct surface measurement independent of transparency, color, or lighting. We present RadarEye (Fig. 2(b)), which combines AoA–ToF beamforming with a physics-informed temporal tracker. The main challenge is multipath: during pouring, the radar receives overlapping reflections from the liquid surface, gripper, source container, and workspace.

II. System Design

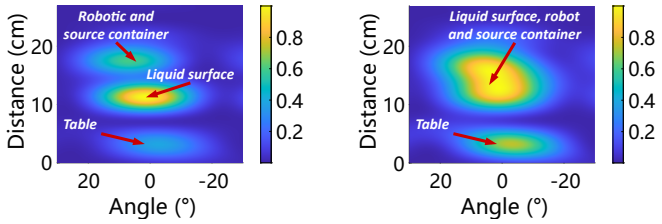
Radar Beamforming for Level Estimation. The radar is placed above the target container. The received signal at the m-th antenna and k-th frequency sample captures all L propagation paths:

$$r_{m,k}(t) = \sum_{l=1}^{L} s(t)\,\alpha_{m,l}(t)\,e^{-j2\pi f_k \tau_l(t)}\,e^{-j2\pi f_k \frac{(m-1)d\cos\theta_l(t)}{c}}, \tag{1}$$

where $\alpha_{m,l}$ is the complex attenuation, f_k the frequency, d the array spacing, c the speed of light, and θ_l/τ_l the AoA/ToF of path l. We discretize the AoA–ToF plane into an $N \times N$ grid and compute signal strength via coherent beamforming: $p_{i,j} = |\mathbf{a}_{i,j}^H \mathbf{r}|$, where $\mathbf{a}_{i,j}$ is the steering vector. This produces the AoA–ToF spectrum $\mathbf{P}$. Under static conditions, the liquid surface produces the dominant peak: $(i^*, j^*) = \arg\max_{(i,j)} \mathbf{P}(i,j)$. During pouring, multipath creates competing reflections (Fig. 3).

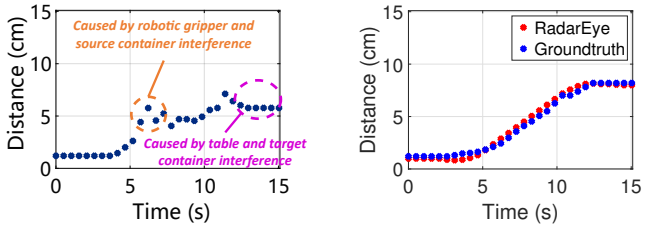
Physics-Informed Liquid Level Tracker. During pouring, multipath from the gripper and source container corrupts peak detection. In the AoA–ToF spectrum, two reflection trajectories appear: one from the liquid surface and one from the gripper/container. To resolve this ambiguity, we jointly optimize bin indices across time. The transition cost from bin (i,j) at time t to (i',j') at $t{+}1$ is:

$$c_t = -\mathbf{P}_t(i,j) - \mathbf{P}_{t+1}(i',j') + \omega\,\eta\big((i,j)\to(i',j')\big), \tag{2}$$



(a) Static: clear dominant peak. (b) Pouring: multipath ambiguity.

Fig. 3. AoA–ToF spectra. During pouring, multipath from the gripper/container overlaps with the liquid surface reflection.



(a) Peak detection fails under multipath. (b) Physics-informed tracker succeeds.

Fig. 4. Tracking comparison. (a) Peak detection jumps between liquid surface and gripper reflections. (b) The proposed tracker follows the liquid surface correctly.

where $\eta = \omega_\theta |i - i'|_2 + \omega_\tau |j - j'|_2$ penalizes large spatial jumps between consecutive frames, since the liquid level changes slowly. The optimal trajectory is:

$$\delta^* = \arg\min_{\delta \in \Delta_T} \sum_{t=0}^{T-1} c_t(\delta(t)), \tag{3}$$

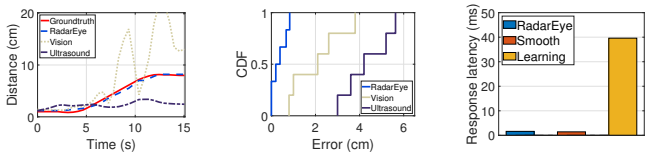
where Δ_T denotes all transition paths up to slot T. Constraining each transition to a Q^2-neighborhood ($Q \ll N$) reduces complexity from $O(N^{2(T-1)})$ to $O(T \cdot Q^2)$, allowing sub-millisecond tracking. Fig. 4 shows that naive peak detection fails under multipath, while the physics-informed tracker correctly follows the liquid surface.

III. Evaluation

Setup. We use a TI IWR6843ISK radar (61.8 GHz, 3.6 GHz bandwidth, 1×4 Tx-Rx) mounted with an Intel RealSense D435i on a Reachy humanoid robot (7-DoF arm). Ground truth is obtained via a ruler inside the container with synchronized video.

Static Level Estimation. In an incremental filling experiment (0–7.4 cm), RadarEye achieves a median error of **0.12 cm**, confirming its ability to localize the liquid surface reflection in the AoA–ToF spectrum.

Dynamic Pouring (Fig. 5). RadarEye achieves **0.35 cm** median error at **0.62 ms** per-update latency, compared with 2.1 cm for the RGB-D camera and 4.3 cm for the ultrasonic sensor. As shown in Fig. 5(a), the radar curve is smooth, the camera curve shows jumps due to missing depth, and the ultrasonic reading is flat because the signal passes through the liquid. Error distributions and response times are in Fig. 5(b) and (c).



(a) Pouring curves. (b) Level estimation errors. (c) Response time.

Fig. 5. RadarEye vs. vision and ultrasound baselines during dynamic pouring.

Tracker Comparison. Compared with envelope smoothing and a deep learning model adapted from [4], the physics-informed tracker matches smoothing speed (∼0.6 ms) with better accuracy under multipath. The deep learning approach requires ∼40 ms per inference.

IV. Conclusion & Next Steps

RadarEye achieves 0.35 cm median error at 0.62 ms latency without training data, outperforming both vision (6×) and ultrasound (12×) baselines. These results confirm that mmWave radar is a practical sensor for liquid level feedback. Next, we plan to integrate RGB and thermal sensors into the fast system and connect the VLM-based slow system for task-level reasoning, toward the full architecture described in Sec. 1.

References

[1] S. Sajjan *et al.*, "ClearGrasp: 3D shape estimation of transparent objects for manipulation", *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[2] H. Deng *et al.*, "FuseGrasp: Radar-camera sensor fusion for transparent object grasping", *IEEE Transactions on Mobile Computing (TMC)*, vol. 24, no. 8, pp 7028–7041, 2025.

[3] H. Lin *et al.*, "PourIt!: Weakly-supervised liquid perception from a single image for visual closed-loop robotic pouring", *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[4] P. Bagad *et al.*, "The sound of water: Inferring physical properties from pouring liquids", *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[5] Y. Hu *et al.*, "Robocap: Robotic classification and precision pouring of diverse liquids and granular media with capacitive sensing", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025.