

Calibrating perceptual trust in robotic vision: The Metacognition-based Adaptive Scanning System (MASS)

Minsung Kang, *University of Georgia, mkang@uga.edu*

Doyeon Lee, *Middle Georgia State University, doyeon.lee@mga.edu*

Abstract

The current paradigm in robotic vision for manipulation is characterized by a fundamental efficiency-accuracy trade-off between low-cost 2D scanning and resource-intensive 3D reconstruction. While constant 3D scanning provides geometric precision, its indiscriminate application leads to excessive computational overhead and temporal delays. Conversely, relying on static 2D perspectives often fails to resolve geometric ambiguities, resulting in grasping failures. To address this dilemma, we propose a Metacognition-based Adaptive Scanning System (MASS) framework. This perspective explores a self-evaluative “Metacognitive Loop” that mimics human cognitive monitoring by quantifying epistemic uncertainty in initial visual data. By dynamically transitioning from 2D to 3D scanning only when information deficiency is detected, the system is conceptually designed to achieve an optimized balance between perceptual reliability and operational efficiency. We analyze this framework through various scenarios of geometric ambiguity, such as altitude deception and orthographic inconsistency, to theoretically conceptualize the necessity of cognitive awareness in autonomous systems. This approach suggests a new direction for resource-efficient, intelligent robotics in safety-critical and dynamic environments.

Keywords: Metacognitive Robotics, Adaptive Scanning, Robot Vision, Uncertainty Quantification, Speed-Accuracy Trade-off, Autonomous Manipulation

Introduction

The rapid advancement of autonomous systems has positioned robotic manipulation as a cornerstone of modern industrial automation and service robotics (Choi et al., 2025). For a robot to interact effectively with its environment, the ability to perceive and interpret complex geometries is paramount (Lashin et al., 2026). Vision-based perception serves as the primary sensory modality, enabling robots to identify, localize, and manipulate objects in diverse settings ranging from structured factory floors to highly unstructured domestic environments (Rady, 2016). As the complexity of these tasks increases, the demand for perception systems that are both computationally efficient and geometrically accurate has become more pressing than ever.

However, a critical “efficiency-accuracy trade-off” persists in the current paradigm of robotic vision (Lin et al., 2024). Modern systems predominantly rely on either high-speed 2D visual sensors or high-fidelity 3D scanning technologies (Gu et al., 2024). While 2D-based systems offer rapid processing suitable for real-time tasks, they are inherently prone to perceptual errors caused by geometric ambiguity, where

different 3D shapes may appear identical from a single orthographic perspective (Hirschmann et al., 2012). Conversely, continuous 3D orbital scanning provides the requisite precision for complex tasks but imposes significant computational overhead and temporal delays, often rendering it impractical for time-sensitive applications (Mihic et al., 2023). This dilemma (see Figure 1 (Chittka et al., 2009)) forces a rigid choice between speed and reliability, often at the expense of operational efficiency. From an information systems perspective, this efficiency-accuracy trade-off directly relates to the challenge of managing computational resource constraints in edge-computing environments. Indiscriminate high-dimensional data acquisition creates processing bottlenecks and latency, which compromises the reliability of safety-critical automated decision-making systems.

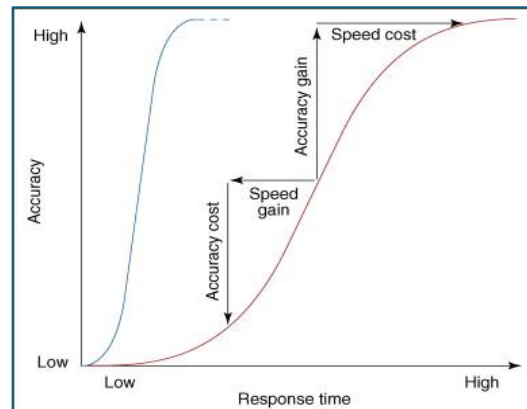


Figure 1. Conceptual illustration of the speed-accuracy trade-off

We argue that the key to transcending this rigid trade-off lies in the integration of metacognitive sensitivity, the computational capacity of a system to evaluate its own perceptual uncertainty in real-time (Abdar et al., 2021; Lee et al., 2025). Rather than treating 2D and 3D scanning as mutually exclusive routines, metacognitive sensitivity functions as an intelligent mediator that “calibrates” the robot's perceptual strategy based on the perceived difficulty of the task. By empowering the robot to recognize its own informational deficiency, it can selectively deploy high-dimensional scanning only when necessary. To explore this potential, this paper introduces a theoretical exploration of the Metacognition-based Adaptive Scanning System (MASS). As a conceptual framework, MASS is designed to offer a computational pathway toward resolving the time-accuracy dilemma by utilizing a self-evaluative loop that calibrates perceptual trust and operational efficiency prior to empirical implementation.

Background and Literature Review

Limitations of fixed perceptual routines in robotic vision

Traditional robotic perception has largely evolved along two divergent paths: high-speed 2D feature extraction and high-fidelity 3D geometric reconstruction. Extensive research in 2D vision has focused on real-time object detection and pose estimation using deep convolutional neural networks (He et al., 2016; Upadhyay et al., 2023). To bridge the gap between these modalities, recent studies have introduced monocular depth estimation and RGB-based height mapping to infer 3D properties from 2D images (Fu et al., 2018; Zhang et al., 2022). However, these software-based estimations are inherently limited by their reliance on learned priors, which often leads to significant “depth-distortion” or “altitude deception” when encountering novel or complex geometries. Unlike physical 3D scanning, which provides ground-truth spatial measurements, RGB-based height estimation lacks the requisite precision for high-stakes manipulation, often resulting in “sloppy” decisions under visual ambiguity (Godard et al., 2019). As

illustrated in Figure 2 (Shen et al., 2025), even a fundamental pick-and-place task requires precise geometric reasoning. (1) The robot must accurately estimate the block's height for a successful 'Pick'. (2) The 'Place' action requires millimeter-level precision to stack the object. Relying solely on high-speed 2D vision can lead to 'altitude deception' and grasp failure, while continuous 3D scanning for such routine tasks significantly reduces operational efficiency.

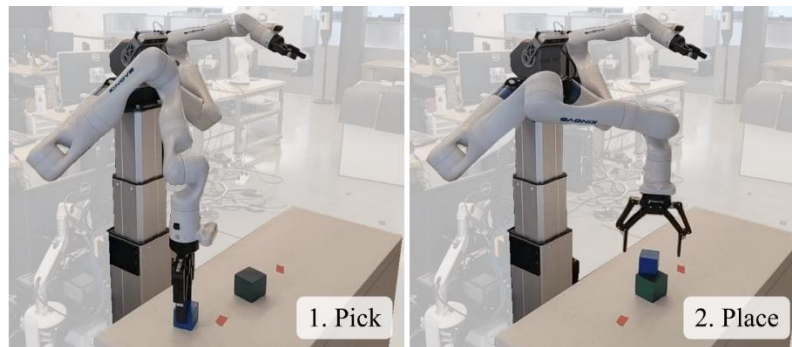


Figure 2. Example of a Pick-and-Place sequence highlighting the perceptual challenge

On the other hand, 3D scanning technologies provide the dense point clouds necessary for resolving such ambiguities. However, the computational cost of continuous 3D data processing remains a significant bottleneck. Studies have shown that indiscriminate 3D scanning during routine tasks introduces unnecessary latency, increasing power consumption and reducing the overall throughput of autonomous systems (Pomerleau et al., 2015). The “time-accuracy dilemma” persists because current systems lack a mechanism to switch between these modalities based on situational necessity, typically relying on a pre-defined, fixed perceptual routine (Bajcsy et al., 2018).

Human-centric active perception and uncertainty

Recent breakthroughs in robotic intelligence emphasize learning from human behavior to resolve the inefficiencies of static perception. Specifically, Xiong et al. (2025) demonstrate that active perception can be effectively learned from human demonstrations, allowing robots to mimic the efficient way humans adjust their viewpoints to gather task-relevant information. This human-centric approach suggests that the key to efficient perception lies in the goal-directed selection of visual data. In this context, uncertainty quantification serves as the internal trigger for such actions. Significant advancements in Bayesian deep learning have enabled neural networks to output not just a prediction, but a measure of “epistemic uncertainty”, representing the model's lack of knowledge regarding a specific input (Hassan & Ismail, 2025).

However, while observing how humans move is crucial, a gap remains in replicating the internal cognitive process that decides when a current perceptual strategy is insufficient. Most active perception frameworks focus on spatial movement (viewpoint planning) rather than the fundamental choice between 2D and 3D modalities. This necessitates a transition from purely behavioral imitation to the integration of metacognitive sensitivity (Lee et al., 2025). By quantifying uncertainty through a metacognitive lens, a robot can determine the exact threshold at which a 2D observation becomes too untrustworthy to proceed, thereby triggering a higher-dimensional scan (Feng et al., 2022).

While these conventional active perception models primarily focus on optimizing spatial positions, determining "where to look" (Bajcsy et al., 2018; Xiong et al., 2025), and uncertainty quantification remains a static statistical byproduct (Hassan & Ismail, 2025), MASS introduces a distinct shift toward modality-shifting active perception. Unlike prior works in sensor fusion (Pomerleau et al., 2015) or robotic self-

evaluation (Feng et al., 2021) that typically rely on fixed sensing routines, MASS operationalizes a real-time metacognitive loop to govern how the system looks, dynamically shifting from 2D to 3D scanning based on situational necessity.

Theoretical framework: Metacognition in autonomous agents

The concept of metacognition, or “cognition about cognition”, offers a robust theoretical solution to the aforementioned challenges by providing a framework for self-reflective decision-making. In cognitive psychology, the foundational model proposed by Nelson (1990) describes metacognition as a dual-level architecture: an object-level (where primary tasks like object recognition occur) and a meta-level (where the system monitors and evaluates those tasks). These levels are linked by two dynamic processes: Monitoring, which carries information from the object-level to the meta-level to assess performance, and Control, which sends instructions back to modify object-level actions based on that assessment (Fleming & Daw, 2017).

A critical metric within this framework is metacognitive sensitivity, which refers to a system's ability to distinguish between its own correct and incorrect judgments (Rahnev & Denison, 2018). In a high-sensitivity system, subjective confidence (perceptual trust) is perfectly aligned with objective accuracy. For autonomous agents, this means the robot should not merely output a result, but also an accurate self-evaluation of how likely that result is to be flawed due to insufficient data. When a robot detects a “metacognitive mismatch,” for instance, recognizing that 2D perception is based on ambiguous visual cues, it can proactively trigger a strategy shift.

The Metacognition-based Adaptive Scanning System (MASS) proposed in this paper mathematically models these psychological principles for an intelligent perception pipeline. By quantifying “information deficiency” through real-time monitoring, MASS calibrates the robot's perceptual trust. Instead of proceeding with a potentially “sloppy” grasp plan based on low-confidence 2D data, the system utilizes the control loop to trigger an adaptive 3D scan. Thus, MASS is structured to transform uncertainty quantification from a mere statistical byproduct into a decisive metacognitive trigger, conceptually guiding the robot to navigate the time-accuracy dilemma with human-like cognitive flexibility.

By embedding this metacognitive monitoring loop, MASS shifts the paradigm from a purely reactive vision script to an adaptive information system. This allows the autonomous agent to recognize when its internal data is insufficient, establishing a foundational architecture for uncertainty-aware decision support and human-AI trust calibration in high-stakes environments.

The MASS Framework

The MASS framework operationalizes the metacognitive monitoring and control loops by quantifying perceptual uncertainty and dynamically adjusting the robot's scanning dimension. This process is structured into three stages: Monitoring, Decision-Making, and Adaptive Action (see Figure 3). To prevent anthropomorphic overstatement, we explicitly separate human subjective cognitive states from the operationalized computational parameters in the MASS framework. Within this engineering pipeline, 'Metacognitive Sensitivity' is defined not as biological introspection, but as the algorithmic capability to differentiate between successful and flawed 2D classifications using statistical metrics. Similarly, 'Confidence' or 'Perceptual Trust' refers strictly to Model Confidence, quantified via calibrated probabilities and calculated structural variances. Specifically, the subjective feeling of doubt is mathematically translated into Epistemic Uncertainty (S_{mt}), while the cognitive shift from fast intuition to deliberate reasoning is operationalized as a threshold-based modality switching trigger mediated by τ .

Stage 1. Metacognitive Monitoring & Uncertainty Quantification

At the object-level, the system acquires an initial 2D top-down image. The meta-level then monitors this perception by computing a composite Metacognitive Score (S_{mt}):

$$S_{mt} = w_1 \cdot U_{semantic} + w_2 \cdot U_{geometric}$$

where w_1 and w_2 are weighting parameters that balance the importance of identity versus physical complexity.

- **Semantic Uncertainty ($U_{semantic}$):** Calculated as the Shannon entropy of the classification softmax output. Higher values indicate class ambiguity.
- **Geometric Uncertainty ($U_{geometric}$):** Quantified by the ratio between the actual contour perimeter and the convex hull length. This captures morphological irregularity.

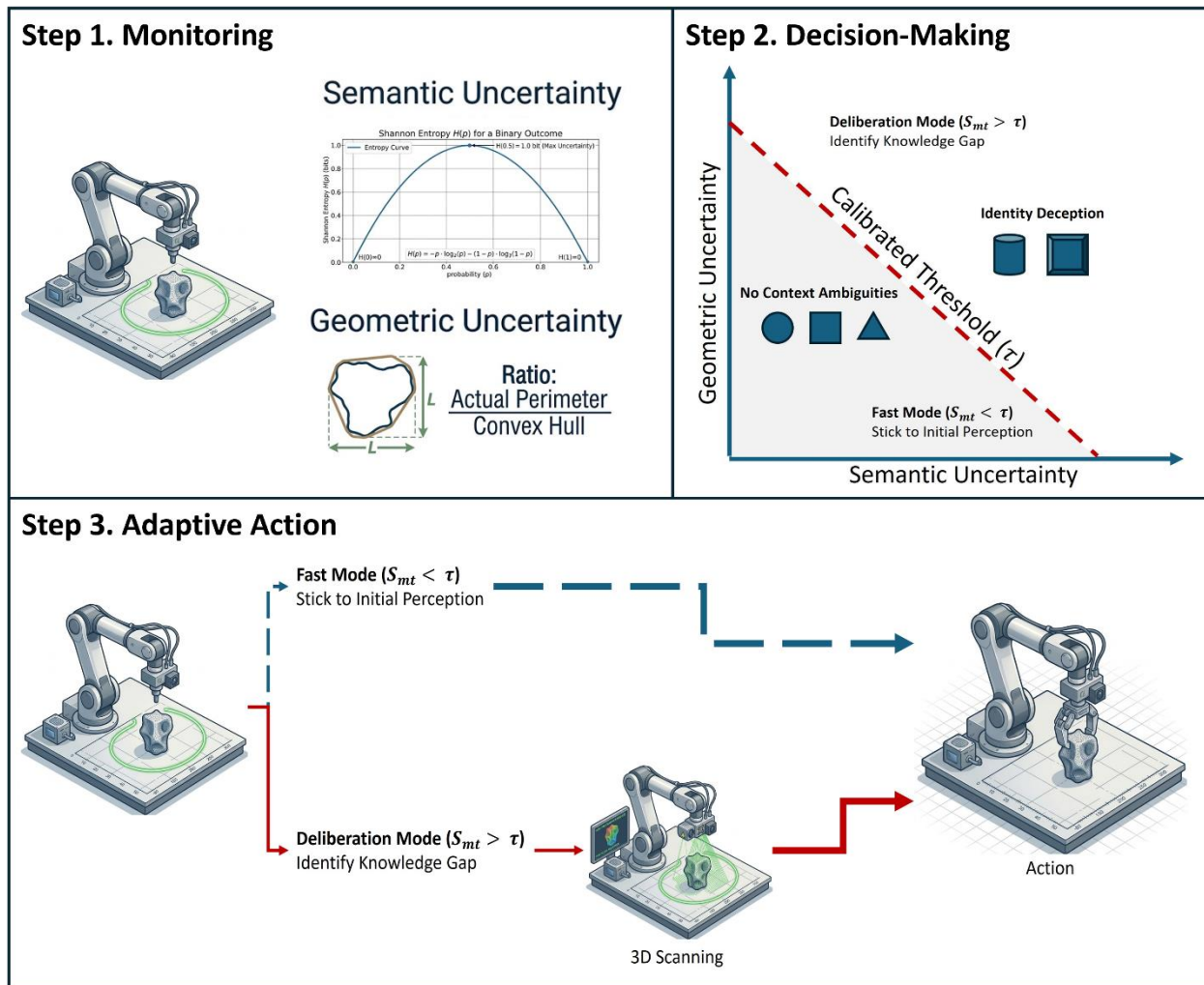


Figure 3. The MASS framework overview

This diagram illustrates the operationalized metacognitive monitoring and control loops, consisting of three stages: (1) Monitoring perceptual uncertainty, (2) Decision-Making for strategic adjustments, and (3) Adaptive Action to dynamically modify scanning dimensions.

In-situ Example (The Identity Deception):

Imagine the robot encounters a circular object from above (see Figure 4). $U_{semantic}$ might be low if the model is 99% sure it is a “Cylinder”. However, if the object is actually a Sphere, the top-down 2D view cannot distinguish the curvature. Here, $U_{geometric}$ acts as a safeguard; by detecting subtle light-shading variances or contour inconsistencies that don't match a perfect flat cylinder, the system raises S_{mt} to signal that the 2D “identity” is a deceptive oversimplification.

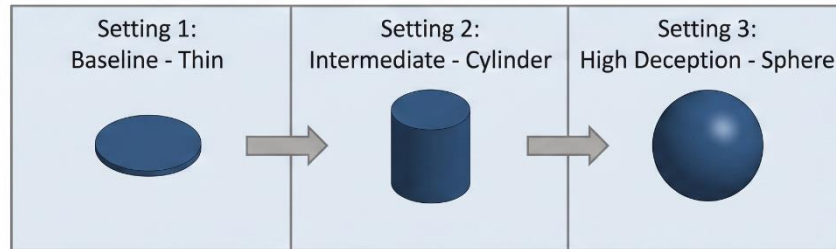


Figure 4. The identity deception case (Altitude Deception)

Conceptual illustration of Scenario 1 (Altitude Deception) where a robot identifies a circular feature from an overhead view. The system's recursive metacognitive loop uses $U_{geometric}$ (detecting subtle shading variances) to cross-check the initial, potentially deceptive 2D classification (Can vs. Sphere), ensuring 3D scanning is triggered to confirm the object's true curvature.

Stage 2. Metacognitive Decision-Making

The control loop determines the scanning strategy by comparing S_{mt} against a dynamically calibrated Metacognitive Threshold (τ):

$$\tau = \mu_{clear} + \kappa \cdot \sigma_{clear}$$

where μ_{clear} and σ_{clear} represent the mean and standard deviation of uncertainty scores obtained from baseline, unambiguous environments. The confidence constant κ (typically 2 or 3) dictates the system's risk tolerance.

- **Cognitive Mode 1: Fast Mode ($S_{mt} < \tau$).** The system concludes the 2D data is sufficient. This “System 1” thinking allows for immediate, high-speed manipulation.
- **Cognitive Mode 2: Deliberation Mode ($S_{mt} > \tau$).** The system identifies a high risk of Altitude or Morphological Deception. It deliberately pauses to gather more information.

In-situ Example (The “T-shaped Block” Trap):

A T-shaped block placed upright might appear as a simple small square from a 2D overhead view (see Figure 5). If the robot proceeds in Fast Mode, it will likely calculate the Center of Mass (CoM) based on that small square, leading to an unstable grasp and a dropped object. However, a high S_{mt} , triggered by the incongruence between the expected weight and the perceived 2D area, forces the system into Deliberation Mode, triggering a 3D scan to reveal the hidden “arms” of the T-block.

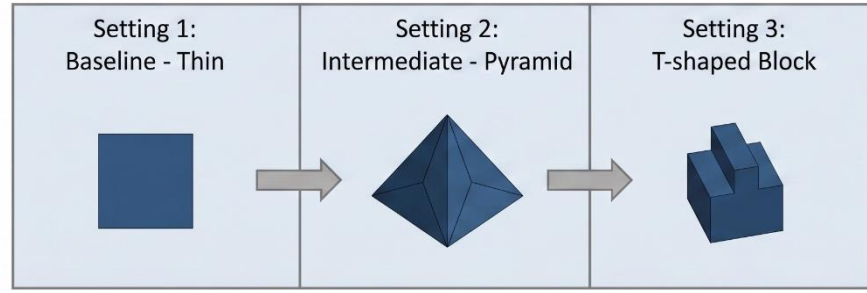


Figure 5. The “T-shaped Block” trap (Morphological Deception)

Visual illustration of Scenario 2 (Morphological Deception). When encountering a complex morphology like an upright T-shaped block, a “metacognitive mismatch” (incongruence between expected weight and perceived 2D area) triggers a high S_{mt} score. This forces the system from fast mode into deliberation mode, utilizing the control loop to activate a 3D scan and resolve the unstable grasp risk.

Stage 3. Adaptive Action and Performance Metrics

Based on the decision in Stage 2, the system executes an Adaptive Action. To theoretically evaluate this hybrid approach in future empirical benchmarks, we propose three primary evaluation metrics:

- **Perceptual Accuracy & Confidence:** Measured by the Task Success Rate (TSR) and Confidence Gain (ΔC), ensuring that the switch to 3D actually leads to a more “trustworthy” state.
- **Operational Efficiency:** Evaluated via Total Cycle Time (T_{total}):

$$T_{total} = T_{scan} + T_{decision} + T_{motion}$$

MASS aims to minimize T_{total} by invoking T_{scan} only when $S_{mt} > \tau$.

- **Metacognitive Reliability:** Defined as the frequency of “True Positives” (correctly switching to 3D for complex shapes like Pyramids or Spheres) versus “True Negatives” (correctly staying in 2D for simple flat objects).

Discussion

The Metacognition-based Adaptive Scanning System (MASS) proposed in this paper introduces a theoretical framework for transitioning from static, reactive perception to a self-reflective, active perceptual paradigm. By operationalizing the “Metacognitive Loop”, this framework is conceptualized to address the long-standing efficiency-accuracy trade-off that has hindered the deployment of robots in complex, real-world environments. In this section, we discuss the broader implications of MASS for computational economy, safety-critical decision-making, and the evolution of general-purpose autonomous agents.

In current robotic paradigms, the “brute-force” application of high-fidelity sensors like 3D LiDAR or structured-light scanners is often viewed as a hallmark of precision. However, from a metacognitive perspective, indiscriminate high-dimensional scanning is a symptom of informational inefficiency. The MASS framework posits that high-dimensional data is not a constant necessity but a conditional asset. By utilizing the integrated uncertainty score (S_{mt}), a robot can reserve its most computationally expensive resources for scenarios that truly demand them, such as resolving the “Altitude Deception” in spheres or the “Morphological Deception” in T-blocks. This selective deployment significantly reduces the total cycle

time (T_{total}) and minimizes energy consumption, which is critical for battery-operated mobile manipulators and edge-computing platforms. The ability of an agent to say, “I see enough to act”, versus “I need more data”, represents a more bio-inspired and resource-efficient approach to intelligence.

One of the most significant contributions of the MASS framework is its role as a “Cognitive Gatekeeper”. In industrial and domestic settings, a single “sloppy” decision, such as grasping an object based on an incorrect 2D height estimate, can lead to collision, damage, or task failure. Traditional systems often proceed with high confidence in low-fidelity data because they lack the internal mechanism to evaluate their own ignorance. MASS explicitly models this “informational deficiency”. By calibrating the metacognitive threshold (τ), we can tune a robot's risk tolerance. In high-stakes environments, such as surgical robotics or nuclear decommissioning, τ can be tightened to ensure the system defaults to deliberation mode at even the slightest geometric ambiguity. This adaptability ensures that the transition to 3D scanning is not merely a technical choice but a strategic decision to maintain safety and reliability under visual uncertainty. Consequently, the implications of MASS extend beyond robotics into Human-Computer Interaction (HCI) and adaptive automation. By ensuring that the system's objective accuracy matches its internal confidence, MASS provides a rigorous framework for designing self-evaluative AI systems that know the boundaries of their own knowledge, a critical requirement for establishing user trust in autonomous enterprise systems.

Current literature on active perception predominantly focuses on “Where to look next”, often referred to as viewpoint planning. While critical, this behavioral focus misses the cognitive dimension of “How to look”. The MASS framework bridges this gap by introducing modality-shifting as a core component of active perception. The transition from 2D orthographic perspectives to 360-degree orbital 3D scans is not just a movement of the end-effector; it is a fundamental update of the agent's belief state. By aligning subjective confidence with objective accuracy, a property of metacognitive sensitivity, MASS ensures that the agent's internal model of the world remains grounded in reality. This framework suggests that true “intelligence” in robotics lies not in the complexity of the sensors, but in the system's ability to evaluate the quality of the information those sensors provide.

While MASS provides a robust conceptual framework, its practical implementation faces several challenges that offer fertile ground for future research. First, the calibration of weighting factors and the threshold (τ) currently relies on pre-calibrated baseline environments. Future work should explore the use of Reinforcement Learning (RL) to allow robots to autonomously learn these thresholds through repeated interaction, effectively developing “situational intuition”.

Furthermore, while this paper focuses on the visual domain, the metacognitive loop could be expanded to multi-modal fusion. For instance, integrating tactile feedback (haptics) as a third uncertainty metric could resolve ambiguities that even a 3D scan might miss, such as material slip or internal weight distribution. Expanding MASS into a multi-sensory metacognitive framework would pave the way for robots that can navigate truly open-ended, unpredictable environments with human-like cognitive flexibility.

Conclusion

The quest for autonomous robotic manipulation has long been caught in a binary struggle between the speed of 2D vision and the precision of 3D scanning. This paper has introduced the Metacognition-based Adaptive Scanning System (MASS), a framework designed to transcend this rigid trade-off by embedding self-evaluative intelligence into the perceptual pipeline. By operationalizing the “Metacognitive Loop”, MASS

moves beyond static routines, allowing robots to proactively navigate visual ambiguity through the quantification of their own uncertainty.

With the critical scenarios, such as Altitude Deception and Morphological Deception, we have argued theoretically that the key to operational efficiency lies not in more powerful sensors, but in more “aware” systems. MASS ensures that high-dimensional, resource-intensive scanning is no longer a default burden but a strategic response to identified informational deficiency. This approach presents a novel architectural solution to the time-accuracy dilemma, prioritizing computational economy without sacrificing the reliability essential for safety-critical tasks.

In conclusion, the MASS framework suggests that the next generation of autonomous agents must do more than just see; they must understand the limits of what they see. By integrating metacognitive sensitivity, we pave the way for robots that possess the cognitive flexibility to adapt to unstructured, dynamic environments with human-like discernment. This shift from “always-on” high-precision scanning to “as-needed” adaptive perception represents a vital step toward truly intelligent, resource-aware robotics capable of navigating the complexities of the real world.

References

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarencov, V., & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76, 243-297.
- Bajcsy, R., Aloimonos, Y., & Tsotsos, J. K. (2018). Revisiting active perception. *Autonomous Robots*, 42(2), 177-196.
- Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed–accuracy tradeoffs in animal decision making. *Trends in ecology & evolution*, 24(7), 400-407.
- Choi, Y. W., Lee, J., Lee, Y., Lee, S., Jeong, W., Lim, D. Y., & Lee, S. W. (2025). A vision-guided adaptive and optimized robotic fabric gripping system for garment manufacturing automation. *Robotics and Computer-Integrated Manufacturing*, 92, 102874.
- Feng, D., Harakeh, A., Waslander, S. L., & Dietmayer, K. (2021). A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 9961-9980.
- Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological review*, 124(1), 91.
- Fu, H., Gong, M., Wang, C., Batmanghelich, K., & Tao, D. (2018). Deep ordinal regression network for monocular depth estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2002-2011).

- Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3828-3838).
- Gu, Q., Kuwajerwala, A., Morin, S., Jatavallabhula, K. M., Sen, B., Agarwal, A., Rivera, C., Paul, W., Elis, K., Chellappa, R., Gan, C., de Melo, C., M., Tenenbaum, J. B., Torralba, A., Shkurti, F., & Paull, L. (2024, May). Conceptgraphs: Open-vocabulary 3d scene graphs for perception and planning. *In 2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5021-5028). IEEE.
- Hassan, M. M., & Ismail, H. R. (2025). Bayesian deep learning applied to diabetic retinopathy with uncertainty quantification. *Heliyon*, 11(2).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hirschmann, M. T., Wagner, C. R., Rasch, H., & Henckel, J. (2012). Standardized volumetric 3D-analysis of SPECT/CT imaging in orthopaedics: overcoming the limitations of qualitative 2D analysis. *BMC medical imaging*, 12(1), 5.
- Lashin, M., El-Mashad, S. Y., & Elgammal, A. T. (2026). Geometry-Guided Traversability Estimation for Compact Indoor Robots via Feature-Level Sensor Fusion. *IEEE Access*.
- Lee, D., Pruitt, J., Zhou, T., Du, J., & Odegaard, B. (2025). Metacognitive sensitivity: The key to calibrating trust and optimal decision making with AI. *PNAS nexus*, 4(5), pgaf133.
- Lin, J., Liu, L., Lu, D., & Jia, K. (2024). Sam-6d: Segment anything model meets zero-shot 6d object pose estimation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 27906-27916).
- Mihić, M., Sigmund, Z., Završki, I., & Butković, L. L. (2023). An analysis of potential uses, limitations and barriers to implementation of 3D scan data for construction management-related use - are the industry and the technical solutions mature enough for adoption?. *Buildings*, 13(5), 1184.
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. *In Psychology of learning and motivation* (Vol. 26, pp. 125-173). Academic Press.
- Pomerleau, F., Colas, F., & Siegwart, R. (2015). A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends® in Robotics*, 4(1), 1-104.
- Rady, S. (2016). Vision-based hybrid map-building and robot localization in unstructured and moderately dynamic environments. *In Novel Applications of Intelligent Systems* (pp. 231-250). Cham: Springer International Publishing.
- Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and brain sciences*, 41, e223.

- Shen, W., Garrett, C., Kumar, N., Goyal, A., Hermans, T., Kaelbling, L. P., Lozano-Perez, T., & Ramos, F. (2025). Differentiable GPU-Parallelized Task and Motion Planning. *Proceedings of Robotics: Science and Systems (RSS) XXI*, Los Angeles, California.
- Upadhyay, R., Asi, A., Nayak, P., Prasad, N., Mishra, D., & Pal, S. K. (2023). Real-time deep learning-based image processing for pose estimation and object localization in autonomous robot applications. *The International Journal of Advanced Manufacturing Technology*, 127(3), 1905-1919.
- Xiong, H., Xu, X., Wu, J., Hou, Y., Bohg, J., & Song, S. (2025). Vision in Action: Learning Active Perception from Human Demonstrations. *Proceedings of The 9th Conference on Robot Learning, in Proceedings of Machine Learning Research*, 305, 5450-5463.
- Zhang, Y., Yi, J., Chen, Y., Dai, Z., Han, F., & Cao, S. (2022). Pose estimation for workpieces in complex stacking industrial scene based on RGB images. *Applied Intelligence*, 52(8), 8757-8769.