# Adaptive Scale-Aware Visual Servoing With Deep Perception and Online Jacobian Learning

Albert Osei Owusu, Paulo Menezes, Rui Cortesão
Electrical and Computer Engineering Department, Institute of Systems and Robotics
University of Coimbra, Portugal
Emails: {albert.owusu, pm, cortesao}@isr.uc.pt

***Abstract*—This work develops an uncalibrated image-based visual servoing framework for an eye-in-hand Meca500 manipulator. The visual state combines the image-plane target centroid coordinates with the logarithm of the largest valid contour area measured inside a YOLO bounding box. This replaces bounding-box width as the scale cue, enabling continuous lateral alignment and monocular approach regulation. A local composite image Jacobian is initialised from central-difference joint perturbations and refined online using gated Broyden updates. Joint velocities are generated by bounded regularised least squares subject to velocity, acceleration, and predicted joint-position limits. A supervisory state machine, stale-frame rejection, contour-validity checks, and communication watchdogs provide safe stopping, grasp verification, and recovery. Ongoing experiments compare centroid-only, fixed-Jacobian scale-aware, and adaptive-Jacobian controllers.**

***Index Terms*—image-based visual servoing, contour area, uncalibrated robotics, online Jacobian learning, constrained control**

## I. Introduction

Conventional preprogrammed manipulation becomes less effective when the target pose or appearance varies. Model-based visual servoing addresses part of this limitation through closed-loop visual feedback, although performance still depends on interaction-matrix accuracy, depth estimates, calibration, and model uncertainty [1]. Such model-based approaches can reduce computational and actuation demands while providing predictable timing, stability, and safety [2]–[4].

However, these methods are less suitable for dynamic or unstructured environments because predefined motion sequences cannot easily adapt to changes in target position, illumination, object appearance, or workspace configuration [5]. Environmental changes may therefore require manual modification and recalibration of the control system, limiting flexibility and scalability [6]. Tasks involving uncertainty, object motion, or human-robot interaction instead require continuous perception and real-time adaptation [7].

This work develops an eye-in-hand uncalibrated image-based visual-servoing framework for robotic manipulation. YOLO localises the target, while the image-plane centroid and logarithm of the largest valid contour area provide position and scale feedback. A local composite image Jacobian is identified directly from joint perturbations and updated online, allowing the robot to regulate target alignment and approach motion without requiring an analytical interaction matrix or explicit hand-eye calibration.

The main contribution is integrating YOLO-based target localisation, largest-valid-contour log-area scale feedback, and uncalibrated local composite Jacobian identification within a single closed-loop controller. Constraint-aware joint-velocity control and task-level safety supervision support reliable hardware operation.

## II. Related Work

Image-based visual servoing (IBVS) enables a robot to regulate its motion directly from visual feedback. Existing approaches can generally be classified as model-based, state-estimation-based, and data-driven methods. Model-based IBVS uses analytical interaction matrices and calibrated geometric models, with recent studies improving robustness to depth variation, calibration errors, and visual-feature uncertainty [8]–[10]. State-estimation approaches incorporate observers, extended Kalman filters, and visual-inertial estimation to recover unknown depth or system states [11]. Although these methods can improve control stability, they remain dependent on accurate models, estimator design, and parameter tuning. Data-driven methods reduce these dependencies by learning visual features, control policies, or mappings between image observations and robot commands [12]–[15].

Recent learning-based IBVS studies demonstrate different strategies for improving adaptability. Fei et al. [16] developed an end-to-end reinforcement-learning framework using deep visual features, while Zeng et al. [17] combined offline and online neural learning to estimate unknown manipulator Jacobians. Wen et al. [18] formulated acceleration-level IBVS as a constrained optimisation problem and introduced an online learned virtual Jacobian. Hao et al. [19] integrated object detection, reference-trajectory generation, and motion prediction for improved tracking. Aracil et al. [20] also investigated model-free Jacobian learning, while more recent approaches use deep reinforcement learning [21] and [22] combined data-driven Jacobian estimation, feature tracking, joint-limit handling, and singularity-aware inversion for redundant manipulators. Despite these advances, reliable real-time operation, stability, and generalisation under changing visual conditions remain important challenges. Therefore, this study adopts a data-driven IBVS framework that learns and updates the visual-to-motion relationship online, reducing reliance on fixed analytical models while supporting accurate and adaptive robot control.

## III. Proposed Framework

The proposed framework uses YOLOv8n to localise the target region, after which contour processing estimates the image-plane centroid and largest valid contour area within the detected region. These visual features generate an error signal for an adaptive image Jacobian and a bounded joint-velocity controller. The Meca500 executes the commands in a closed loop, while a task supervisor manages alignment, approach, grasping, lifting, and safety stops. Figure 1 shows the complete perception, control, and supervision loop.

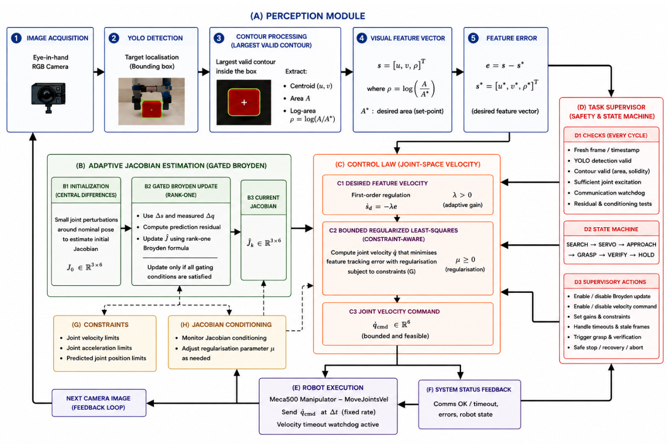


Fig. 1. Proposed scale-aware uncalibrated IBVS pipeline. YOLO-based target localisation and largest-valid-contour processing provide centroid and log-area visual features. A locally identified composite image Jacobian is refined using gated Broyden updates, while bounded regularised least-squares optimization generates constraint-aware joint-velocity commands under task-level safety supervision.

## IV. Implementation and Evaluation

The ongoing evaluation uses an eye-in-hand camera, a six-axis Meca500, and a YOLO target detector. Table I separates the effects of contour-area feedback from those of online model adaptation. Trials vary in lateral target position, starting distance, illumination, and brief occlusion. Primary metrics are final centroid and log-area errors, convergence time, success rate, joint-command smoothness, Jacobian condition number, accepted/rejected adaptation updates, and watchdog interventions. Ablations remove contour gates, gain scheduling, and command smoothing. Results will be reported over repeated randomised trials using medians, interquartile ranges, and 95% confidence intervals.

TABLE I
Controller Variants for the Ablation Study

| Controller | Centroid | Log-area | Online $\hat{\mathbf{J}}$ adaptation |
|---|---|---|---|
| C1: Centroid-only | ✓ | – | – |
| C2: Log-area + fixed $\hat{\mathbf{J}}$ | ✓ | ✓ | – |
| C3: Log-area + adaptive $\hat{\mathbf{J}}$ | ✓ | ✓ | ✓ |

## V. Conclusion

This framework combines YOLO detection, scale feedback based on the largest valid contour area, online composite-Jacobian learning, bounded regularised joint-velocity control, and safety-oriented task supervision. The next stage is a repeatable hardware comparison against fixed-Jacobian and centroid-only baselines, followed by analysis of convergence, robustness, and per-cycle computation time.

## References


[1] Mecademic, *Meca500 Robot User Manual*, 2025, available from Mecademic Robotics.

[2] J. K. Gupta, K. Menda, Z. Manchester, and M. J. Kochenderfer, "Structured mechanical models for robot learning and control," in *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, ser. Proceedings of Machine Learning Research, A. M. Bayen, A. Jadbabaie, G. J. Pappas, P. A. Parrilo, B. Recht, C. Tomlin, and M. N. Zeilinger, Eds., vol. 120. PMLR, Jun. 2020, pp. 328–337.

[3] J. Muru and A. Rassolkin, "A scoping review of energy consumption in industrial robotics," *Machines*, vol. 13, no. 7, Art. no. 542, Jun. 2025.

[4] B. El Mabsout, "Survey on safe robot control via learning," arXiv:2501.01432 [cs.RO], Dec. 2024.

[5] F. Aghili and K. Parsa, "An adaptive vision system for guidance of a robotic manipulator to capture a tumbling satellite with unknown dynamics," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, Sep. 2008, pp. 3064–3071.

[6] O. Ochoa, E. Mendez, C. Lucas-Dophe, J. A. Luna-Sanchez, V. H. Soto-Herrera, D. Olivera-Guzman, M. Alvarado Perez, E. Lugo-del Real, I. N. Ayala-Garcia, and A. Gonzalez, "Flexible human-robot interaction: Collaborative robot integrated with hand tracking," in *2023 XXV Robotics Mexican Congress (COMRob)*, Xalapa/Veracruz, Mexico, Nov. 2023, pp. 25–30.

[7] C. Lynch and P. Sermanet, "Language conditioned imitation learning over unstructured data," in *Robotics: Science and Systems XVII*, D. A. Shell, M. Toussaint, and M. A. Hsieh, Eds., Virtual Event, Jul. 2021.

[8] Q. Gongye, P. Cheng, and J. Dong, "Image-based visual servoing with depth estimation," *Transactions of the Institute of Measurement and Control*, vol. 44, no. 9, pp. 1811–1823, 2022.

[9] C. Zhang, Z. Xie, and G. Chen, "Key-points based visual servo for accurate cross-task robot manipulations," in *Proceedings of the 2024 International Conference on Big Data Mining and Information Processing*, ser. BDMIP '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 151–157.

[10] Z. Zhang, Y. Yang, W. Zuo, G. Song, A. Song, and Y. Shi, "Image-based visual servoing for enhanced cooperation of dual-arm manipulation," *IEEE Robotics and Automation Letters*, vol. 10, no. 4, pp. 3374–3381, Apr. 2025.

[11] F. Xu and P. Chen, "Research on uncalibrated adaptive visual servoing control based on dual-camera fusion," *Advances in Mechanical Engineering*, vol. 17, no. 9, pp. 1–14, 2025.

[12] X. Zhong, Q. Zhou, Y. Sun, S. Kang, and H. Hu, "Deep reinforcement learning-based uncalibrated visual servoing control of manipulators with FOV constraints," *Applied Sciences*, vol. 15, no. 8, Art. no. 4447, 2025.

[13] S. Auddy, A. Paolillo, J. Piater, and M. Saveriano, "Imitation learning-based direct visual servoing using the large projection formulation," *Robotics and Autonomous Systems*, vol. 190, Art. no. 104971, 2025.

[14] P. Katara, H. YVS, H. Pandya, A. Gupta, A. Sanchawala, G. Kumar, B. Bhowmick, and M. Krishna, "DeepMPCVS: Deep model predictive control for visual servoing," in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, 2021, pp. 2006–2015.

[15] S. Felton, E. Fromont, and E. Marchand, "Deep metric learning for visual servoing: When pose and image meet in latent space," in *2023 IEEE International Conference on Robotics and Automation*, London, United Kingdom, May 2023, pp. 741–747.

[16] H. Fei, Z. Wang, and A. Kennedy, "Robust reinforcement learning-based visual servoing with convolutional features," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 9781–9786, 2023.

[17] H. Zeng, Y. Lyu, J. Qi, S. Zou, T. Qin, and W. Qin, "Adaptive finite-time model estimation and control for manipulator visual servoing using sliding mode control and neural networks," *Advanced Robotics*, vol. 37, no. 9, pp. 576–590, 2023.

[18] L. Wen and Z. Xie, "A data-driven acceleration-level scheme for image-based visual servoing of manipulators with unknown structure," *Frontiers in Neurorobotics*, vol. 18, Art. no. 1380430, 2024.

[19] Z. Hao, D. Zhang, and B. Honarvar Shakibaei Asli, "Motion prediction and object detection for image-based visual servoing systems using deep learning," *Electronics*, vol. 13, no. 17, Art. no. 3487, 2024.

[20] R. Aracil, N. Garcia, C. Perez, L. Paya, J. M. Sabater, J. M. Azorin, and L. M. Jimenez, "Robust image-based visual servoing system using a redundant architecture," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 445–450, 2005.

[21] Z. Li, Y. Zhou, L. Wang, X. Zhang, A. Li, M. Zhu, and Q. Wu, "An end-to-end controller with image-based visual servoing of industrial manipulators with soft actor-critic algorithm," *Knowledge-Based Systems*, vol. 311, Art. no. 112980, 2025.

[22] Z. Xie, Y. Zheng, and L. Jin, "A data-driven image-based visual servoing scheme for redundant manipulators with unknown structure and singularity solution," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 10, pp. 6230–6241, Oct. 2024.