KFUPM
Journal of Undergraduate Research International

Diamond Open Access

Article
Control and Instrumentation Engineering

Vision-Guided Cooperative Manipulation Framework for Synchronized Grasping and Transport of Cylindrical Objects Using Dual Mobile Manipulators

Muteb Hojayj Aljohani, Mohammed Abdel-Nasser and Muhammad Faizan Mysorewala*

Cite https://doi.org/10.64589/juri/231576

Submitted: April 15, 2026 **Revised:** May 8, 2026 **Accepted:** May 15, 2026

ABSTRACT

Despite advances in cooperative mobile manipulation, reports on low-cost frameworks that enable dual-robot grasping and transport without specialized hardware remain limited. This study introduces and experimentally validates a vision-guided, synchronized cooperative manipulation framework for two DJI RoboMaster EP mobile manipulators performing a cylindrical object pick-and-place task in warehouse-relevant settings. The key contribution of this work is an event-driven synchronization mechanism for dual-robot cooperative manipulation, which enables coordinated grasping and transport without direct robot-to-robot wireless messaging, depth sensors, external motion-capture infrastructure, or pre-built 3D maps. The manipulation pipeline includes visual lateral alignment, visual forward approach, grasp execution, synchronized lifting, cooperative transport, and coordinated release. Visual guidance is achieved using color-based HSV detection and proportional control, providing a lightweight perception and control solution for cooperative object handling. The framework is evaluated through fixed-position repeatability and position generalization studies. In a five-trial repeatability study, the system achieves a 100% task success rate with a mean completion time of 65.26 ±1.65 s and coefficient of variation of 2.5%. In a six-configuration generalization study, the system achieves 100% success, with task completion times ranging from 50.8 s to 86.5 s. Across all eleven trials, the proposed scheme maintains safe and correct task progression. Grip synchronization latency increases with asymmetry in the robots' approach distances, with symmetric and asymmetric configurations producing latencies ≤ 6 s and ≥ 22 s, respectively. Overall, event-driven synchronization offers a practical, lightweight coordination mechanism for cooperative dual-robot manipulation under controlled laboratory conditions, with minimal sensing and communication requirements.

Keywords: cooperative manipulation, event-driven synchronization, mobile manipulators, multi-robot systems, visual servoing, warehouse automation

1. INTRODUCTION

Mobile manipulation is increasingly recognized as a practical strategy for robots to operate beyond structured factory settings by combining navigation with physical interaction [1,2]. However, as deployments expand to larger objects, single-robot manipulation becomes limiting: many objects in logistics, maintenance, and service robotics — such as long cylindrical pipes, tubes, and rods — are elongated, heavy, or difficult to stabilize with a single platform [3–5]. Their geometry introduces contact uncertainty, rolling susceptibility, and strong coupling between grasp quality and base motion [5,6], making reliable handling challenging even with well-resolved perception [1–3]. Cooperative manipulation, in which multiple robots share perception, grasping, and transport [7,8], offers a promising solution, enlarging the set of manipulable objects while reducing per-robot mechanical demands. However, such schemes introduce challenges [9,10]: accurate inter-robot and object alignment, synchronization of grasp/lift events to prevent twisting or slipping, and coordinated transport that maintains stability under disturbances [11].

Recent work has explored multi-agent coordination from learning perspectives, including decentralized agents exchanging compact intent representations [12,13] and structured priors for sample-efficient transfer [14,15]. Despite this progress, two gaps persist for repeatable, low-cost cooperative transport. First, many approaches assume rich sensing (external tracking, precise object-pose estimation) or focus on policy learning, obscuring failure modes and complicating deployment on resource-limited platforms [11,14,16]. Second, effective cooperative transport methods often rely on carefully tuned inter-robot controllers, compliant hardware, or specialized grippers [7,9], limiting accessibility for rapid prototyping and educational use [3]. Bridging robust cylindrical-object perception [15,17] into a complete cooperative pipeline on commodity manipulators thus remains an open challenge.

To address these gaps, this study introduces a practical cooperative grasp-and-transport framework implemented and validated on two RoboMaster EP Core manipulators. The system uses a lightweight perception–action loop: visual side-alignment followed by targeted approach toward the cylindrical object via color-segmentation and centroiding for visual servoing. Once each robot reaches its terminal approach condition, the manipulators execute coordinated grasp poses and close their grippers. Cooperative manipulation is enforced through an explicit synchronization protocol ensuring both robots complete gripping before initiating shared lift-and-transport motion, concluding with coordinated release.

The primary contributions of this work can be summarized as follows:

- Development of a complete vision-guided cooperative manipulation pipeline for dual mobile manipulators, including alignment, approach, grasping, synchronized lifting, transport, and release of a cylindrical object.
- Formulation of a lightweight single-host event/barrier-based synchronization strategy, enabling coordinated grasp confirmation and cooperative transport without direct robot-to-robot wireless messaging.
- Experimental validation on dual DJI RoboMaster EP platforms, demonstrating repeatability, positional generalization, and quantitative coordination performance.

The remainder of this paper is organized as follows. Section 2 reviews related work on cooperative mobile manipulation and positions the proposed approach within the existing literature. Section 3 describes the system, including the hardware platform, the event-driven synchronization architecture, and the task execution phases. Section 4 presents the experimental results from the repeatability and generalization studies. Section 5 concludes the paper and outlines directions for future work.

2. RELATED WORK

Existing research on cooperative mobile manipulation spans three themes: (i) coordination and transport controllers for multi-robot systems, (ii) perception and task decomposition for mobile manipulation, and (iii) object-specific grasping pipelines, particularly for cylindrical geometries. The present work sits at the intersection of these themes, targeting a repeatable, low-cost cooperative transport framework with an end-to-end perception–action pipeline and explicit synchronization on commodity mobile manipulators.

2.1. Cooperative Transport and Multi-Robot Coordination. Cooperative transport has been addressed via constrained optimization, including hierarchical quadratic programming for kinematic and cooperative constraints[4] and distributed optimization for scalable, decentralized computation[7]; both typically assume reliable state feedback and careful tuning under perception noise and contact uncertainty. Formation-based approaches instead regulate inter-robot spacing and orientation through hierarchical constraint-based control[9], improving transport stability but degrading when grasp success or contact is uncertain — motivating explicit state-machine and synchronization logic. Learning-based coordination, such as cooperative reinforcement learning for multi-robot tasks[11], can yield adaptive policies but demands substantial training data and sim-to-real effort, often obscuring failure causes in physical experiments. Applied systems integrating perception with real-time execution under dynamic, obstacle-dense conditions[3] further illustrate that practical deployment still depends on rich sensing and extensive engineering for synchronization and recovery.

2.2. Mobile Manipulation Pipelines and Task Decomposition. Structured task decomposition is widely used to reduce planning complexity, e.g., target-driven multi-subscene manipulation that structures the environment into sub-scenes[8], though at the cost of added system complexity and reliance on accurate entity perception. Learning-based priors, such as reachability behavior priors guiding base-pose selection[16], improve single-robot task success but do not address the coupling constraints introduced by cooperative transport. Minimal-perception approaches, including pure object detection without full pose reconstruction[10], are attractive for low-cost platforms but are less effective when contact geometry (e.g., long cylinders) or multi-agent synchronization is required.

2.3. Perception for Cylindrical Objects and Real-Time Grasping. Cylindrical and box-like objects have motivated specialized perception pipelines. Elliptical shape primitives enable real-time manipulation with strong interpretability under motion[18], but require consistent contour visibility and can degrade under occlusion or clutter. Combining geometric primitives with depth information supports reliable grasp-pose estimation in human–robot collaboration[19], though the required depth-sensing stack may be unavailable on low-cost platforms. Learning-based detection and localization on point-cloud or stereo/depth data improves generalization across objects and scenes[20], at the cost of higher computational load, data requirements, and the continued need for explicit execution logic to handle grasp confirmation and multi-agent synchronization.

2.4. Multi-Robot Cooperative Grasp-and-Transport Systems. Practical systems often improve contact robustness through mechanical compliance: Tejada et al.[21] combined FinRay-inspired soft grippers with caging/closure-based control for two-robot transport, though relying on specialized end-effectors and planar caging conditions. Wu *et al.*[22] proposed an intention-communication mechanism in which agents exchange spatial intention maps under learned policies, validated via sim-to-real foraging tasks rather than synchronized dual-ended grasping of a single long object. Learning-based shape detection has also improved real-time grasping of cylindrical and cubic objects under perception noise[13], though with added dataset dependence and validation overhead when transferred to new platforms or lighting conditions.

2.5. Positioning of the Proposed Work. A recurring gap across these directions is the lack of an end-to-end cooperative pipeline that is both lightweight enough for commodity manipulators without external tracking, and engineered with explicit synchronization and logging for reproducible experimentation. This work addresses that gap using a pragmatic cooperative transport baseline on two RoboMaster EP Core manipulators, leveraging onboard RGB cues for alignment and approach and explicit synchronization for grasp confirmation and transport onset — prioritizing interpretability, repeatability, and transparent execution. Unlike communication-based approaches that rely on direct inter-robot messaging[7,9], the proposed event/barrier-based scheme avoids the latency, packet-loss risk, and implementation complexity of robot-to-robot communication, trading full adaptability for simplicity and reliability suited to tasks where phase sequencing can be defined a priori.

Table S1 in the Supplementary Information shows that representative systems typically rely on external infrastructure (e.g., motion capture, fiducials) or single-robot assumptions, whereas the proposed system targets low-cost dual-robot grasp-and-transport using onboard RGB cues and explicit grasp synchronization.

3. SYSTEM DESCRIPTION

3.1. Hardware Platform. The experimental platform consists of two DJI RoboMaster EP robots (DJI Innovations, Shenzhen, China), each equipped with a four-wheel Mecanum chassis[16], a two-joint robotic arm, a pneumatic gripper, an onboard RGB camera (1280×720 px), programmable LED armor panels, and a Wi-Fi module for station-mode connectivity. Robot control runs via the RoboMaster Python SDK on a Huawei laptop (Huawei Technologies Co., Ltd., Shenzhen, China) serving as the single host computer. The target object is a 1.2-m-long white hollow cylinder resting on a box at a consistent gripping height, with a blue stripe at each end (Phase 1 cue) and a red rim (Phase 2 reference).

3.2. System Architecture. The system follows a single-host, event/barrier-based architecture: two independent software threads on one host computer each manage one robot through seven sequential sub-phases — (1) blue lateral alignment, (2) red rim approach, (3A) synchronized gripping, (3B) cooperative lifting, (4) sideways transport, (5A) coordinated placement, and (5B) gripper release. Three bilateral barriers, implemented via shared Python threading.Event flags (grip, lift, placement), require each robot to set its own flag upon completing a phase and wait for its partner's flag before proceeding. Safety timeouts (300 s for grip/lift; 180 s for placement) prevent indefinite blocking. Fig. 1 illustrates the per-robot execution state machine.

3.3. Task Execution.

3.3.1. Phase 1: blue lateral alignment. Each robot performs lateral alignment by detecting the blue stripe on its assigned cylinder end. The camera stream is processed via OpenCV HSV thresholding ($H \in [90°, 130°]$, $S,V \geq 80$), comparing the largest blue contour's horizontal centroid against the frame center (640 px). The robot translates at 0.07 m/s until blue is absent for eight consecutive frames. The full procedure is detailed in Algorithm S1 (Supplementary Information).

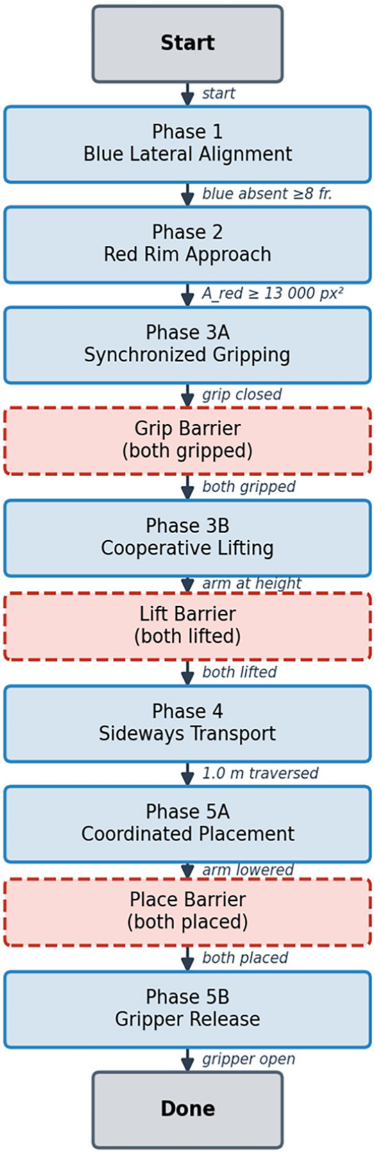

Figure 1. Per-robot execution state machine. Red dashed boxes indicate bilateral synchronization barriers: each robot places a block until both threads signal completion

3.3.2. Phase 2: red rim approach. Each robot then approaches its cylinder end using the red rim as reference. A dual-range HSV mask ($H \in [0°, 10°] \cup [160°, 180°]$) robustly detects red[15], with a proportional yaw controller $\omega = K_p \dot{\Delta} x$ ($K_p = 0.003$ rad px^{-1} s^{-1}). Approach terminates when the red-rim area exceeds 13,000 px^2(~15.8 FPS control loop). The full procedure is given in Algorithm S2 (Supplementary Information); resulting centering-error profiles are shown in Fig. 2.

3.3.3. Phases 3A–5B. In Phase 3A (Synchronized Gripping), each robot moves its arm to the grasp pose $(x_{v}g, y_{v}g)$, closes its gripper to 60%, sets its gripped flag, and waits at the grip barrier; grasp confirmation here refers to successful completion of the gripper-close command, not physical contact verification through force, tactile, or visual feedback. In Phase 3B (Cooperative Lifting), each robot raises its arm to the defined lift height. In Phase 4 (Sideways Transport), both robots translate 1.0 m

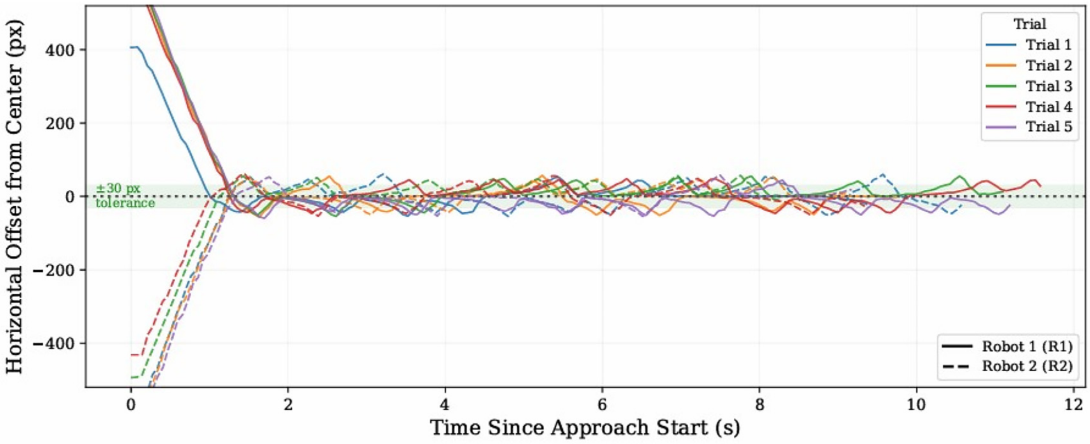

Figure 2. Horizontal centering error of both robots over time during Phase 2 (red rim approach) across five trials. The yaw controller of each robot adjusts orientation based on the centroid offset ($cx - 640$), driving the horizontal offset toward zero. Dotted green lines indicate the ±30 px tolerance band, within which the robot is considered aligned to the red-rim center. Solid and dashed lines correspond to Robot 1 (R1) and Robot 2 (R2), respectively

at 0.1 m/s[16]. In Phase 5A (Coordinated Placement), each robot lowers the object to the placement height, pauses 1.5 s, and sets its placed flag at the placement barrier. In Phase 5B, both robots open their grippers to 50%. This full procedure is summarized in Algorithm S3 (Supplementary Information).

3.4. LED Phase Signaling.

The LED armor panels of each robot display a phase-specific color: blue (Phase 1), red (Phase 2), green (Phase 4), or yellow (Phase 5A). The panels are turned off in Phase 5B.

3.5. System Implementation Parameters.

Table S2 in the Supplementary Information summarizes the key implementation parameters used across all system phases, including vision and HSV detection settings, motion control values, arm and gripper configurations, and barrier-timeout thresholds.

4. EXPERIMENTAL RESULTS

4.1. Experimental Setup.

All experiments were conducted on a flat tile floor under consistent fluorescent lighting. The manipulated object was a 1.2-m-long white hollow cylinder, tested across two campaigns: five fixed-position trials (repeatability) and six initial-position configurations (generalization). Fig. 3 shows all seven task phases from a representative trial, from Phase 1 (blue lateral alignment) to Phase 5B (gripper release), with LED colors confirming each active phase.

4.2. Fixed-Position Repeatability Study.

All five repeatability trials were completed successfully (100%); Table 1 summarizes per-trial outcomes and statistics. Mean completion time was 65.26 ±1.65 s ($CV = 2.5\%$); grip synchronization latency ranged from 1.00–4.13 s, mean 2.63 ±1.36 s ($CV \approx 52\%$). The higher CV in grip latency reflects natural trial-to-trial variation in Phase 2 duration, which depends on real-time visual

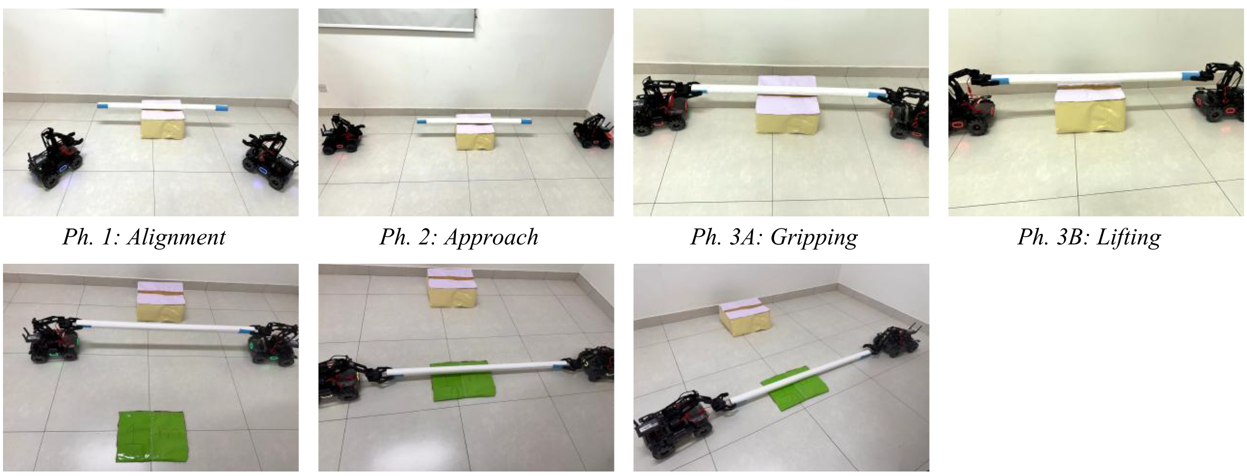

Figure 3. Images of seven task phases and sub-phases from a representative trial. Row 1: Phases 1–3B. Row 2: Phases 4–5B. LED colors confirm each phase: blue (Phase 1), red (Phase 2), green (Phase 4), yellow (Phase 5A), and off (Phase 5B)

Table 1. *Fixed-position repeatability trial outcomes and statistical summary ($n = 5$)*

Trial	Time to Both Gripped (s)	Grip Latency (s)	Task Time (s)
1	36.74	3.96	64.55
2	37.78	1.00	65.02
3	36.75	4.13	63.71
4	38.78	2.03	68.02
5	38.74	2.02	65.00
Mean (5/5)	**37.76**	**2.63**	**65.26**
SD	1.01	1.36	1.65
Min	36.74	1.00	63.71
Max	38.78	4.13	68.02
CV (%)	2.7	51.7	2.5

feedback and control timing rather than a fixed schedule; this does not compromise safety, as the barrier architecture ensures Phase 3B is never initiated until both robots confirm gripping. Fig. 4 presents representative camera frames from R1 during Phases 1–2, showing the centroid coordinates and offsets guiding each phase. Fig. 5 presents the phase-duration breakdown for both robots, highlighting consistent timing across trials. Fig. 6 shows the chassis X–Y trajectories of both robots, where closely overlapping paths confirm reliable, repeatable routes.

4.3. Generalization to Variable Initial Positions.

All six generalization trials were completed successfully (100%) across the range of starting configurations tested; Table 2 summarizes the initial positions and per-trial metrics. Trials

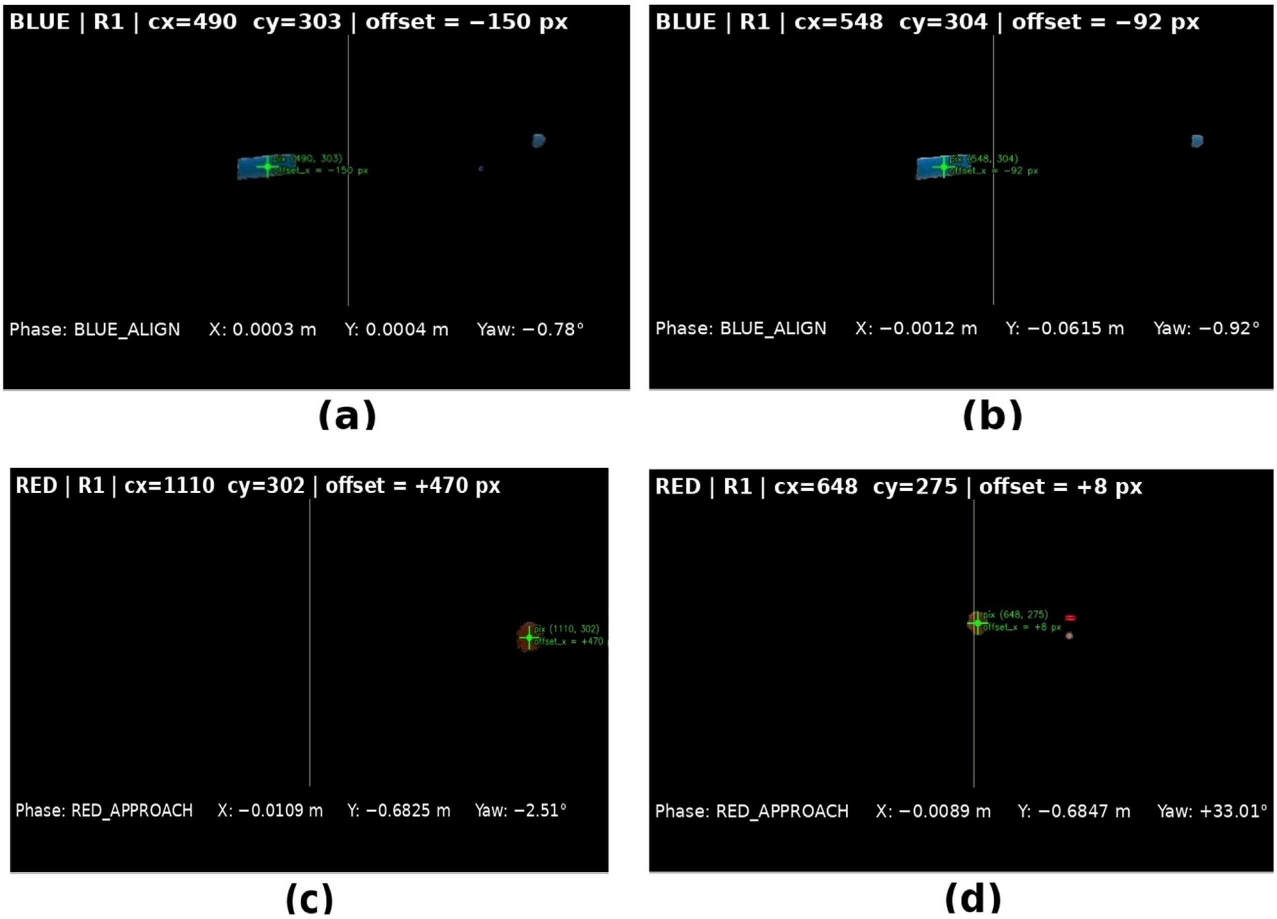

Figure 4. Camera frames captured by Robot 1 (R1) across two operational phases: (a) and (b) correspond to Phase 1 (blue lateral alignment), while (c) and (d) pertain to Phase 2 (red rim approach). Each frame shows the x, y, and yaw coordinates of the detected centroid location, along with the centroid axis coordinates cx and cy relative to the center of the camera frame. These values represent the lateral offset or error used to determine the direction of lateral movement in Phase 1 or the orientation adjustment required in Phase 2

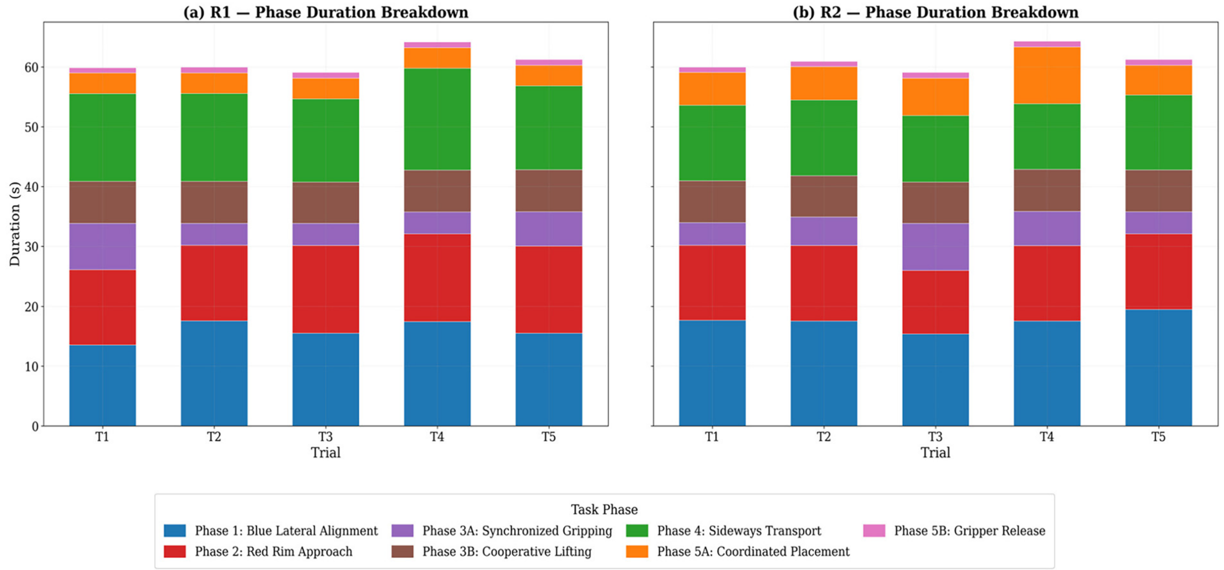

Figure 5. Phase-duration breakdown across five repeatability trials: (a) Robot 1, (b) Robot 2

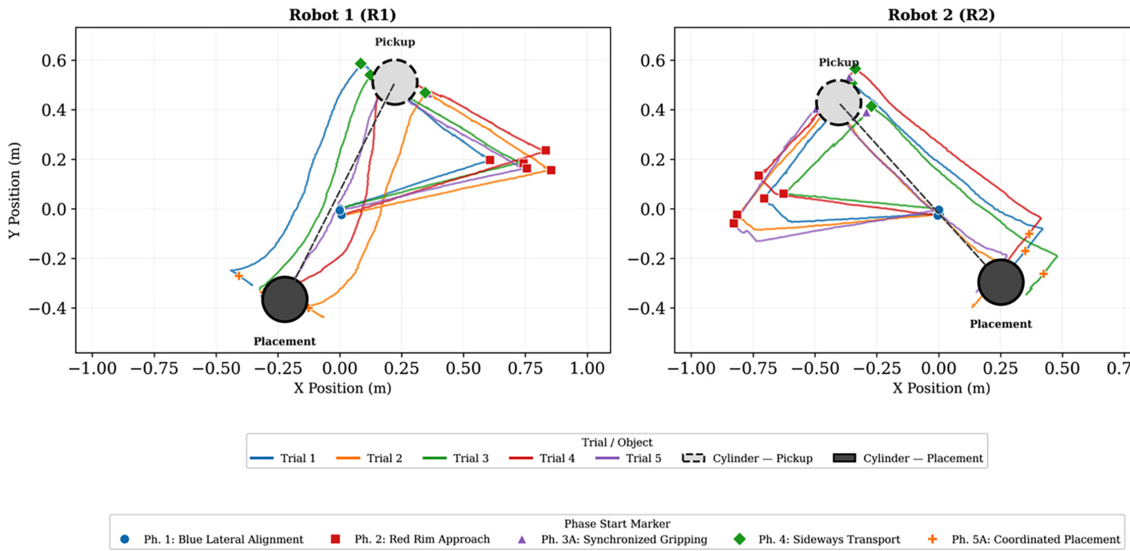

Figure 6. Chassis X–Y trajectories for Robot 1 (R1) and Robot 2 (R2) across five repeatability trials. Paths were reconstructed from chassis odometry data (position subscriptions at 10 Hz via the RoboMaster SDK), with the origin defined at each robot's starting position at the beginning of Phase 1. Consistent paths confirm reliable transport execution

Table 2. *Initial Position Configurations — Real Trial Results (one trial each)*

Trial	Configuration	Type	$d(\mathrm{R1})$ (m)	$d(\mathrm{R2})$ (m)	$\lvert\Delta d\rvert$(m)	Task Time (s)
T6	R1 Far, R2 Close	Asym.	1.980	0.537	1.443	83.02
T7	Both Close	Sym.	0.663	0.577	0.086	62.77
T8	Both Far	Sym.	1.296	1.775	0.479	78.94
T9	R1 Close, R2 Far	Asym.	0.374	1.801	1.427	76.47
T10	Both Very Close	Sym.	0.173	0.251	0.078	50.78
T11	Both Very Far	Sym.	1.479	2.119	0.640	86.51
Mean	**All 6**	—	—	—	—	**73.08**

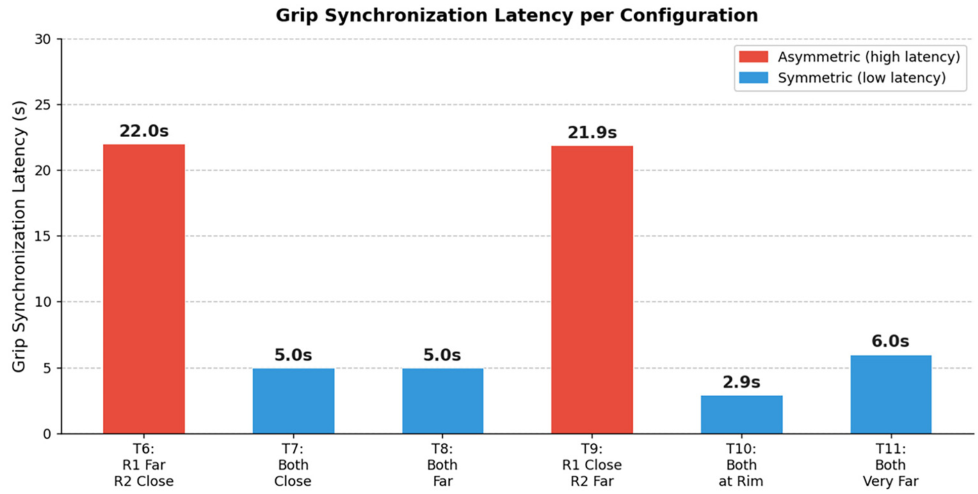

Figure 7. Grip synchronization latency for each generalization configuration (T6–T11). Red bars indicate asymmetric configurations ($|\Delta d| \geq 1.4$ m), which produced latencies of approximately 22 s. Blue bars indicate symmetric configurations ($|\Delta d| \leq 0.64$ m), which produced latencies below 6 s

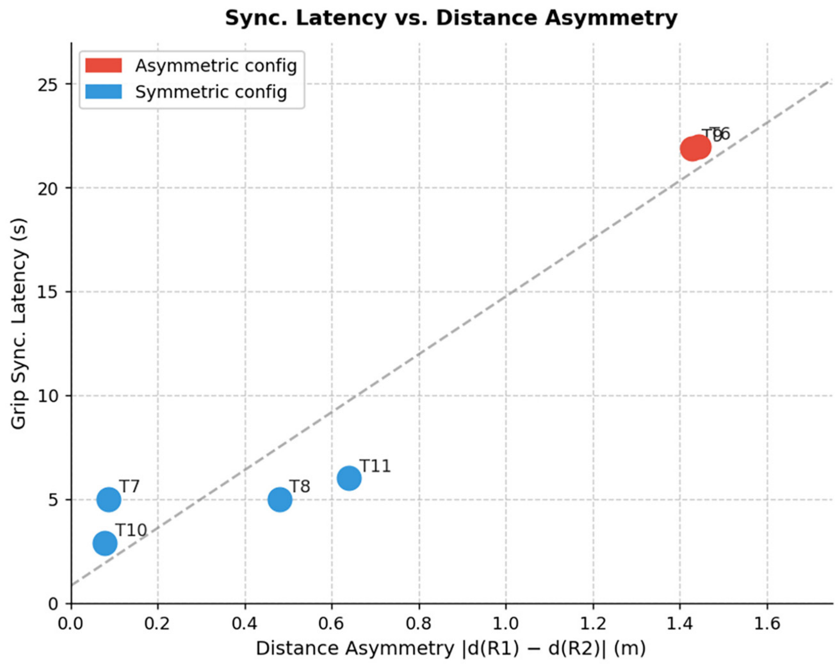

Figure 8. Grip synchronization latency as a function of approach distance asymmetry $|d(R1)-d(R2)|$ across six generalization configurations. The dashed trend line indicates a positive correlation between asymmetry and latency. Red and blue markers denote asymmetric and symmetric configurations, respectively

covered symmetric configurations (similar approach distances) and asymmetric configurations (one robot considerably closer than the other); the asymmetry $|\Delta d|$ ranged from 0.078 m (T10, both at rim) to 1.443 m (T6, R1 far/R2 close). Task completion times ranged from 50.78–86.51 s, with longer times generally corresponding to farther starting positions, and a mean of 73.08 s across all six trials.

Fig. 7 presents grip synchronization latency per configuration: asymmetric configurations (T6, T9) produced latencies of approximately 22 s, versus below 6 s for symmetric configurations. Fig. 8 confirms a positive correlation between latency and $|\Delta d|$ across all six configurations.

5. CONCLUSION

This work demonstrates that synchronized cooperative manipulation of a cylindrical object can be achieved using two mobile manipulators with lightweight onboard vision and event-driven coordination, supporting repeatable task execution across tested configurations and confirming the effectiveness of simple visual feedback and explicit synchronization for structured dual-robot manipulation. Coordination efficiency depends on the robots' relative approach conditions: greater approach-distance asymmetry increases synchronization delay, indicating that balanced approach paths improve task efficiency while maintaining safe, successful cooperative execution.

The current implementation is limited by its dependence on tuned HSV thresholds, fixed transport conditions, and a single-host coordination architecture. Future work must thus be aimed at improving robustness and flexibility through adaptive approach-speed modulation; enhancing perception under varying lighting; online trajectory replanning; depth-sensor integration; testing under different object geometries, weights, and surface properties; validation in more realistic warehouse conditions; quantitative comparison with communication-based cooperative manipulation methods; and introduction of more distributed coordination architectures.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Muteb Hojayj Aljohani – *Department of Control and Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0009-0005-3434-4041
Email: s202155690@kfupm.edu.sa

Mohammed Abdel-Nasser – *Research Collaborator, Department of Control and Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0009-0003-0288-4348
Email: g202203800@kfupm.edu.sa

Corresponding Author

Muhammad Faizan Mysorewala – *Research Mentor, Associate Professor, Department of Control and Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*; 0000-0001-8909-8937
Email: mysorewala@kfupm.edu.sa

ACKNOWLEDGMENTS

The authors acknowledge the laboratory and institutional support that made this work possible. Experimental trials were conducted from February to April 2026 as part of a semester-long project within the RES201 (Vertically Integrated Research) course.

REFERENCES

(1) D. Honerkamp et al. Language-grounded dynamic scene graphs for interactive object search with mobile manipulation. *IEEE Robot. Autom. Lett.* **9**, 8298–8305 (2024).

(2) A. Cardenas et al. Vision-based control of a mobile manipulator with an adaptable-passive suspension for unstructured environments. *J. Mech. Robot.* **13**, 050908 (2021).

(3) A. A. Adil et al. A multi-robot collaborative manipulation framework for dynamic and obstacle-dense environments. *Front. Robot. AI* **12**, 1585544 (2025).

(4) D. Koung et al. Cooperative multi-robot object transportation system based on hierarchical quadratic programming. *IEEE Robot. Autom. Lett.* **6**, 6466–6472 (2021).

(5) Z. Pan et al. Algorithms and systems for manipulating multiple objects. *IEEE Trans. Robot.* **39**, 2–20 (2022).

(6) K. Oishi & T. Jimbo. "Autonomous cooperative transportation system involving multi-aerial robots with variable attachment mechanism," in Proc. IEEE/RSJ IROS" 2021, pp. 6322–6328.

(7) X. Meng et al. A discrete-time distributed optimization algorithm for cooperative transportation of multi-robot system. *Complex Intell. Syst.* **10**, 343–355 (2024).

(8) J. Hou et al. "TaMMa: Target-driven multi-subscene mobile manipulation," in 8th Annu. Conf. Robot. Learn." 2024.

(9) X. Li et al. Hierarchical constraint-based formation control strategy for multi-robot cooperative transportation system in warehousing scenarios. *J. Vib. Control*, 10775463251389027 (2025).

(10) B. Griffin. "Mobile robot manipulation using pure object detection," in Proc. IEEE/CVF WACV" 2023, pp. 561–571.

(11) J. Baltes, I. Akbar & S. Saeedvand. Cooperative dual-actor proximal policy optimization algorithm for multi-robot complex control task. *Adv. Eng. Inform.* **63**, 102960 (2025).

(12) J. Wu et al. "Spatial intention maps for multi-agent mobile manipulation," in Proc. IEEE ICRA" 2021, pp. 8749–8756.

(13) H. Dong, J. Zhou & H. Yu. Robotic grasps of cylindrical and cubic objects via real-time learning-based shape detection. *IEEE Trans. Autom. Sci. Eng.* **22**, 9681–9697 (2024).

(14) F. Schmalstieg et al. Learning hierarchical interactive multi-object search for mobile manipulation. *IEEE Robot. Autom. Lett.* **8**, 8549–8556 (2023).

(15) H. Gu et al. Mean-field multiagent reinforcement learning: A decentralized network approach. *Math. Oper. Res.* **50**, 506–536 (2025).

(16) S. Jauhri, J. Peters & G. Chalvatzaki. Robot learning of mobile manipulation with reachability behavior priors. *IEEE Robot. Autom. Lett.* **7**, 8399–8406 (2022).

(17) E. Sihite et al. Multi-modal mobility morphobot (M4) with appendage repurposing for locomotion plasticity enhancement. *Nat. Commun.* **14**, 3323 (2023).

(18) H. Dong et al. Real-time robotic manipulation of cylindrical objects in dynamic scenarios through elliptic shape primitives. *IEEE Trans. Robot.* **35**, 95–113 (2018).

(19) H. Dong, J. Zhou & H. Yu. Real-time human-robot collaborative manipulations of cylindrical and cubic objects via geometric primitives and depth information. *arXiv:2106.14461*, (2021).

(20) Z. Zhou et al. Learning-based object detection and localization for a mobile robot manipulator in SME production. *Robot. Comput.-Integr. Manuf.* **73**, 102229 (2022).

(21) J. C. Tejada et al. Enhancing object manipulation and transportation in multi-robot systems with soft gripper integration and caging-based control. *J. Intell. Robot. Syst.* **111**, 76 (2025).