Commun Transp Res

Perspective

From sky to doorstep: Heterogeneous embodied transportation for the urban last mile

Ruo Chen[1], Jiacheng Wu[2], Fei Gao[1,3], Yang Zhu[1,4,✉]

Cite this article: https://doi.org/10.26599/COMMTR.2026.9640057

1 Introduction

The most difficult part of the urban last mile is often not the final kilometer but the final few hundred meters. The final segment crosses incompatible access conditions: road vehicles may stop short of building entrances, curbs and stairs interrupt wheeled routes, and aerial routes require safe landing and handover sites. These gaps span airspace, roads, sidewalks, communities, and buildings.

Urban aerial mobility is developing through an ecosystem that includes electric vertical take-off and landing vehicles, unmanned aircraft, automation, and supporting institutions (Cohen et al., 2021). This ecosystem perspective motivates an aerial transport layer, yet an aerial segment cannot complete door-to-door service unless it connects with ground circulation, pedestrian infrastructure, and building access.

The growing autonomy of unmanned aerial vehicles (UAVs) and ground robots creates an opportunity to coordinate different mobility modes across these access gaps, assigning each route segment to a suitable embodiment.

Operational research has established selected pairings: truck–UAV models jointly optimize routing and scheduling for parcel delivery (Murray and Chu, 2015). A broader survey of cooperative heterogeneous multi-robot systems identifies task decomposition, task allocation, perception, multi-agent planning, and control as connected challenges, and calls for end-to-end autonomous solutions to be tested in less restrictive smart-city settings (Rizk et al., 2019). These studies motivate cross-platform coordination, but they do not themselves establish an end-to-end UAV–wheeled–legged delivery chain. Three transportation questions therefore remain for the perspective developed here: which part of a task should be assigned to each embodiment, where and how cargo should be transferred, and how safety should be managed when robotic traffic enters human-centered shared spaces. Fig. 1 provides a visual roadmap by linking the urban delivery setting, the complementary roles of aerial, wheeled, and legged robots, and the coordination and safety mechanisms required to operate them as one transport system.

2 Capability complementarity

We use heterogeneous embodied transportation to describe a mobility system in which autonomous agents with different physical forms jointly serve different segments of an end-to-end transport task. Its central principle is embodiment–environment fit: the body of a transport agent determines the surfaces, obstacles, interfaces, and social spaces it can negotiate. Fig. 1(b) maps operating domains to robot modes: flight connects separated urban areas, wheels serve structured ground routes and precise handovers, and legs extend access across uneven terrain. Quadrupedal locomotion has demonstrated traversal over deformable terrain, rubble, and vegetation (Lee et al., 2020), although parcel access near buildings remains a proposed application rather than an established delivery result. A robot is therefore treated here not merely as a device carrying a parcel, but as a transport mode with a specific operating domain, infrastructure dependence, risk profile, and service cost.

Fig. 1(a) places the three modes in a common urban setting without prescribing a fixed handover sequence. One possible delivery chain uses an aerial segment to bypass congested ground links, transfers the parcel at an approved site, and continues along structured ground paths. A legged robot could take over where stairs or damaged pavement prevent further wheeled access. Depending on local conditions, the route could instead use a direct UAV-to-legged transfer, terminate at a smart locker, or remain on a single platform. The mode sequence should therefore follow access conditions and transfer costs.

This perspective changes the optimization objective. A heterogeneous system must account not only for travel time and energy, but also for transfer delay, synchronization uncertainty, infrastructure occupancy, safety exposure, failure recovery, and fleet repositioning. A mode switch is justified only when the accessibility or efficiency gained from changing embodiment exceeds the transfer penalty. Heterogeneity is therefore valuable when a route contains spatial discontinuities that a single platform would otherwise traverse slowly, unsafely, or not at all.

3 Task division, transfer, and closed-loop coordination

3.1. Dynamic task division

Multi-robot allocation ranges from independent assignments to coupled, multi-step tasks. Because task decomposition, allocation, perception, planning, and control interact in heterogeneous teams (Rizk et al., 2019), dynamic task division is a core mechanism in Fig. 1(c). Assignments should account for shipment urgency and payload, energy reserves, weather, surface conditions, pedestrian density,

Just Accepted

[1] College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China. [2] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore. [3] Huzhou Institute of Zhejiang University, Huzhou 313000, China. [4] Ningbo Innovation Center, Zhejiang University, Ningbo 315100, China.

✉ Corresponding author. E-mail: zhuyang88@zju.edu.cn

Received: September 7, 2026; Revised: September 28, 2026; Accepted: October 7, 2026

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, http://creativecommons.org/licenses/by/4.0/).

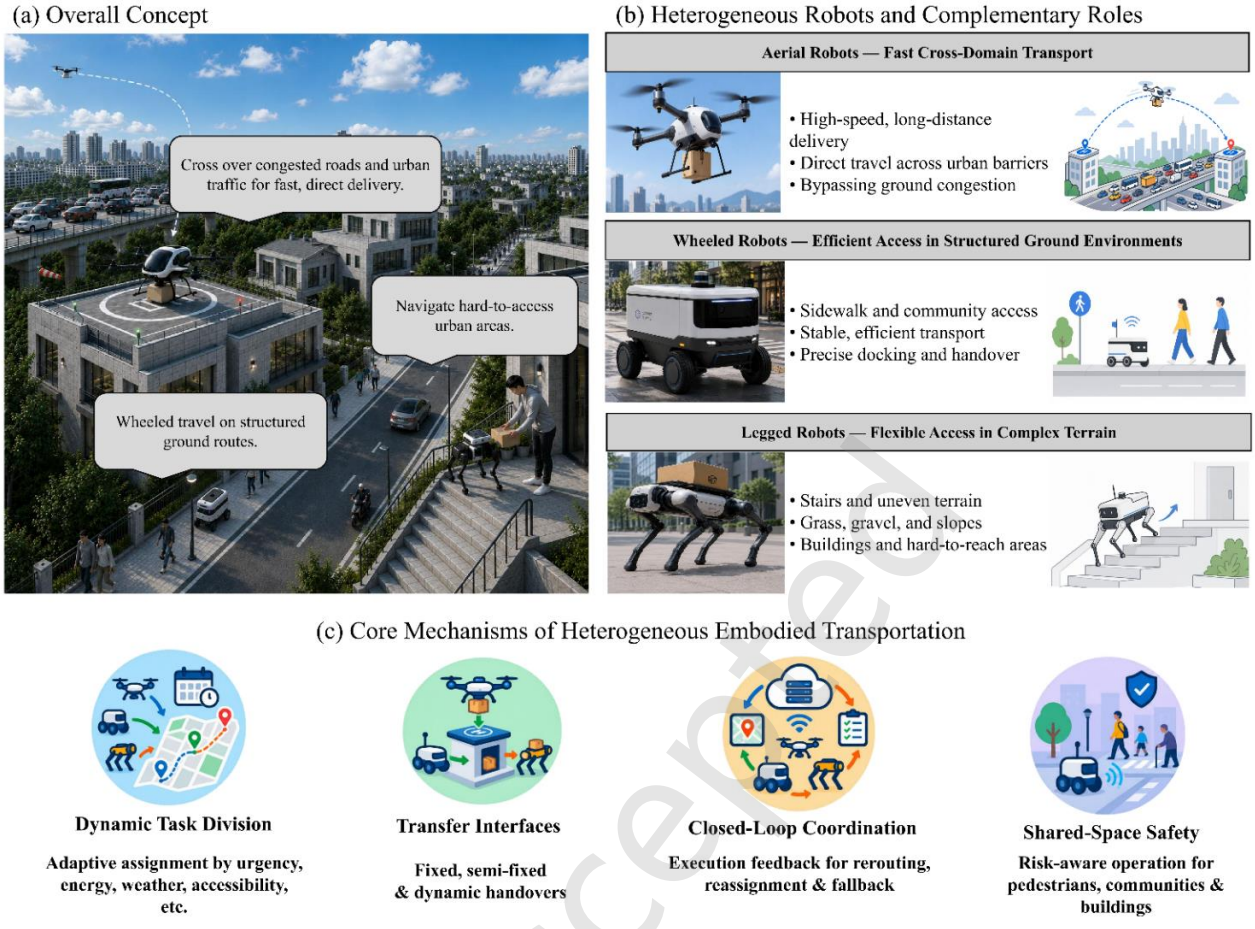


Fig. 1. Conceptual framework for heterogeneous embodied transportation in the urban last mile. (a) An illustrative urban setting with aerial travel over congested roads, wheeled travel on structured ground routes, and legged access at stairs and building approaches. (b) Complementary operating domains and proposed roles of aerial, wheeled, and legged robots. (c) Four mechanisms for integration: dynamic task division; fixed, semi-fixed, and dynamic transfer interfaces; closed-loop coordination; and shared-space safety.

airspace restrictions, transfer-node availability, building access, and failure consequences. Mode selection then weighs the time saved by aerial travel, the efficiency and payload stability of rolling on structured routes, and the access gained through legged locomotion against transfer costs. Legged robots are most useful where their terrain capability offsets their lower speed or endurance.

Assignments must also adapt as conditions change after dispatch. Rain, wind, snow, temporary road works, elevator outages, crowding, or a closed entrance may alter the preferred chain. A route planned as UAV–wheeled robot may need to become UAV–legged robot, or a direct handover may need to be replaced by a smart locker or human-assisted delivery. Such adjustments require timely information from vehicles, infrastructure, and the surrounding environment. Reassignment should consider not only route feasibility but also delivery time, energy availability, transfer cost, and safety constraints. Local disruptions may trigger changes to only part of the delivery chain rather than requiring complete replanning. The operational problem is therefore to select and continuously revise the sequence of embodiments that completes the service.

3.2. Transfer interfaces

Transfer interfaces connect the robot roles shown in Fig. 1(b). As indicated in Fig. 1(c), handovers can be fixed, semi-fixed, or dynamic. Fixed transfers occur at purpose-built facilities such as rooftops, neighborhood microhubs, parcel lockers, or building entrances. They support standardized docking, charging, storage, and clear responsibility, but require infrastructure investment and may force detours. Semi-fixed transfers select one of several approved zones according to current demand and local conditions. Dynamic transfers allow two mobile agents to determine a rendezvous point in real time.

Pair-specific delivery studies already demonstrate that aerial and ground operations cannot be scheduled independently. The seminal truck–UAV formulation jointly optimizes vehicle routes, drone sorties, and delivery timing (Murray and Chu, 2015). Extending this operational coupling, an adaptive rendezvous in the proposed framework may reduce waiting or avoid a crowded or obstructed site, but its net value depends on a coupled space–time decision. Both robots must arrive with sufficient energy, the location must remain safe, and a delay by one agent must not induce unsafe hovering or sidewalk blockage. Transfer policies should therefore specify waiting thresholds, fallback nodes, reassignment rules, and responsibility when a handover fails.

Interoperability is equally important. Mechanical interoperability concerns cargo containers, locking devices, docking geometry, and tolerance to localization error. Digital interoperability concerns identity, authentication, task state, maps, and communication protocols. Operational interoperability concerns reservation, priority, charging, maintenance, and liability. Without common interfaces, a nominally multimodal network can fragment into proprietary systems that cannot

Just Accepted

share hubs or exchange tasks.

3.3. Closed-loop coordination

A purely open-loop workflow would assign the transport chain, route each platform, and expect onboard controllers to execute the plan. In the proposed framework, such a workflow is treated as vulnerable to accumulated deviations between network-level assumptions and local execution. The closed-loop mechanism in Fig. 1(c) instead returns execution feedback to routing, reassignment, and fallback decisions. This feedback is especially important in heterogeneous transport, where execution uncertainty can propagate across routes and transfers.

A practical architecture can contain three interacting levels. The operational level allocates orders, selects the embodiment sequence, and reserves network and transfer resources. The transfer-coordination level chooses rendezvous points, synchronizes arrivals, and manages local queues. The execution level controls flight, rolling, walking, docking, and cargo exchange. Feedback should include estimated arrival time, remaining energy, weather exposure, surface traversability, pedestrian density, access-control status, elevator availability, and transfer success. The upper levels can then change a rendezvous point, replace a robot, repartition a task, delay entry into a crowded area, or direct the shipment to a safe fallback facility.

Within this proposal, coordination is distributed between onboard autonomy and infrastructure or cloud support. Local intelligence handles time-sensitive safety responses, while external sensing and computing can supply information for fleet-level decisions. Rooftop sensors, community cameras, building systems, digital maps, and cloud scheduling may provide information unavailable to an individual robot. Because this division of responsibility has not been validated for the complete transport chain considered here, the network should be designed to enter a minimal-risk state when communication or infrastructure support is interrupted.

4 Safety in shared urban spaces

Heterogeneous transport extends robot access from low-altitude airspace and roads to pedestrian spaces and building interiors. This access should not disrupt how people move, live, and work. Safety should therefore guide task allocation, transfers, and execution in Fig. 1(c), rather than remain limited to onboard collision avoidance.

Safety measures should reflect each platform's risks. UAV risks include uncontrolled descent, dropped cargo, and rotor contact. Ground robots may skid, tip over, or lose balance on slopes and stairs. Platforms need reliable perception and localization, health monitoring, secure loads, and safe responses to faults, such as controlled landing or safe stopping. Transfers add risks of misalignment, locking failure, and early cargo release. They require suitable sites and confirmation that the receiving platform is ready before cargo release. Clear procedures are also needed to stop and resume transfers after failures or pedestrian entry.

Shared-space safety also requires predictable behavior and unobstructed access. Building on socially aware navigation research (Singamaneni et al., 2024), evaluation should cover comfort, accessibility, and perceived control, not just collision avoidance. Robots should follow clear speed and yielding rules, signal their intent, and avoid passing too close or forcing pedestrians aside. Tactile paving, curb ramps, and emergency exits should remain clear. Transfer capacity and waiting areas should be planned to prevent queues from blocking entrances and corridors. These measures should account for the needs of children, older adults, and people with disabilities.

Deployment also requires clear access rules, responsibilities across operators, and data protection. Indoor operations should follow building access, property management, and fire-safety requirements. Residents should be able to refuse robot entry. Data collection and storage should be limited to task needs. Urban air mobility research also identifies institutional, regulatory, and public-acceptance challenges (Cohen et al., 2021). Assessment should therefore consider both the benefits to recipients and the effects on non-users, including noise, visual intrusion, and public-space use.

These requirements should shape the choice of routes, platform combinations, and transfer sites. Longer routes or slower delivery may be preferable when they avoid crowds or risky transfers. When indoor operation is unsuitable, lobby delivery may be more appropriate than doorstep delivery. The goal is not to maximize robot reach, but to improve last-mile access while protecting people's safety and use of shared spaces.

5 Deployment pathways and research priorities

Heterogeneous embodied transport should not be presented as a universal replacement for vans, couriers, or single-mode robots. It is most defensible where three conditions hold: the delivery chain contains a meaningful accessibility discontinuity; the benefit of changing embodiment exceeds the transfer and coordination cost; and shared-space risks can be managed through infrastructure and operating rules. Likely early applications include campuses, hospitals, logistics parks, industrial sites, managed residential communities, mountainous neighborhoods, and emergency operations. These environments combine bounded demand, identifiable access points, and stronger control over infrastructure.

A conservative deployment sequence for the proposed concept would begin with fixed transfer nodes and clearly separated operating zones. A later stage could introduce multiple approved transfer points, mixed pedestrian traffic, and limited building integration. Open urban deployment would require interoperable hubs, shared digital infrastructure, cross-operator coordination, standardized building access, and risk-based regulation across air and ground domains. The urban air-mobility ecosystem provides one part of this foundation (Cohen et al., 2021), but the network must continue downward through landing sites, sidewalks, community interfaces, and building access points.

The need for end-to-end heterogeneous multi-robot solutions tested in realistic smart-city settings has already been identified (Rizk et al., 2019), while urban air-mobility research emphasizes that institutions and public acceptance must develop alongside technology (Cohen et al., 2021). For the present concept, those gaps translate into several research priorities. Network design should jointly locate aerial corridors, transfer hubs, ground robot paths, charging facilities, and building interfaces. Operations research should integrate routing, scheduling, fleet repositioning, energy management, and transfer reliability under stochastic demand and environmental conditions. Robotics research should standardize cargo and docking interfaces and quantify the practical capability boundaries of wheeled and legged

Just Accepted

systems. Human-factors research should establish shared space safety metrics, especially for vulnerable users. Governance research should address access rights, liability, cybersecurity, privacy, noise, public-space allocation, and distributional equity. Evaluation must move beyond isolated platform success rates toward end-to-end measures of service completion, generalized cost, failure recovery, human exposure, and accessibility gained.

6 Conclusions

This editorial has framed the final few hundred meters of urban transport as a sequence of discontinuities between airspace, roads, sidewalks, communities, and buildings. It proposes complementary roles for UAVs, wheeled robots, and legged robots, while treating dynamic task division, transfer interfaces, closed-loop coordination, and shared-space safety as unresolved system-design problems rather than validated capabilities. Heterogeneous robots should therefore be investigated as connected transport modes within an integrated accessibility network. The central research question is not how far each robot can travel alone, but under what conditions different embodiments can cooperate safely and efficiently from the sky to the doorstep.

Author contributions

Ruo Chen: Conceptualization, investigation, visualization, writing–original draft. **Jiacheng Wu:** Investigation. **Fei Gao:** Supervision. **Yang Zhu:** Conceptualization, supervision, writing–review and editing. All authors reviewed and approved the manuscript.

Replication and data sharing

No original data or code were generated for this Editorial. All information supporting the discussion is available in the cited literature.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Declaration of generative AI and AI-assisted technologies in the writing process (if applicable)

During the preparation of this work, the authors used OpenAI Codex solely for language polishing and proofreading. The authors reviewed and edited the output and take full responsibility for the content of the manuscript.

References

Cohen, A.P., Shaheen, S.A., Farrar, E.M., 2021. Urban air mobility: History, ecosystem, market potential, and challenges. IEEE Transactions on Intelligent Transportation Systems 22, 6074–6087. doi:10.1109/TITS.2021.3082767.

Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., Hutter, M., 2020. Learning quadrupedal locomotion over challenging terrain. Science Robotics 5, eabc5986. 1126/scirobotics.abc5986, doi:10.1126/scirobotics.abc5986.

Murray, C.C., Chu, A.G., 2015. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. Transportation Research Part C: Emerging Technologies 54, 86–109. doi:10.1016/j.trc.2015.03.005.

Rizk, Y., Awad, M., Tunstel, E.W., 2019. Cooperative heterogeneous multi-robot systems: A survey. ACM Comput. Surv. 52. doi:10.1145/3303848.

Singamaneni, P.T., Bachiller-Burgos, P., Manso, L.J., Garrell, A., Sanfeliu, A., Spalanzani, A., Alami, R., 2024. A survey on socially aware robot navigation: Taxonomy and future challenges. The International Journal of Robotics Research 43, 1533–1572. doi:10.1177/02783649241230562.

Author biography



Ruo Chen received the B.S. degree in control engineering and automation from Ningbo University, Ningbo, China, in 2023. He is currently pursuing the Ph.D. degree with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. His research interests include world models, embodied intelligence, diffusion models, and intelligent robotic systems.



Jiacheng Wu received the M.S. degree in electrical engineering from Anhui University of Technology, Ma'anshan, China, in 2023. From 2025 to 2026, he is a joint doctoral student with Nanyang Technological University, Singapore. His research interests include hybrid systems, reinforcement learning, and robust control.



Fei Gao received the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, China, in 2019. He is currently a tenured professor with the College of Control Science and Engineering, Zhejiang University. His research interests include aerial robots, autonomous navigation, motion planning, optimization, localization, and mapping.



Yang Zhu received the bachelor's degree from Hangzhou Dianzi University, Hangzhou, China, in 2010, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2015. Following research appointments at the University of California San Diego, the University of Texas at Dallas, Nanyang Technological University, and Tel Aviv University, he joined Zhejiang University in 2021. His research interests include control, optimization, learning, intelligent transportation systems, and fuel-cell systems. He has served as an Associate Editor for *Automatica* since 2023.

Just Accepted