

Thematic Section: Vision-based Navigation in Mobile Robotics

In this thematic section of *Computación y Sistemas* (CyS), we present a selection of 10 research papers showcasing recent advances in intelligent mobile robotics through innovations in autonomous navigation, visual perception, motion planning, and AI-driven decision-making. A prominent theme is the growing reliance on vision as a primary sensing modality, particularly for UAVs operating in complex and dynamic environments. The section highlights both learning-based and classical methodologies, demonstrating how deep learning, generative AI, geometric reasoning, and traditional planning techniques can complement one another to achieve robust autonomy. Contributions span emergency-response UAVs, autonomous drone racing, humanoid and biped robot navigation, intelligent transportation systems, and machine-learning-enabled perception, reflecting the multidisciplinary nature of modern mobile robotics. Together, these works emphasize practical deployment, real-time performance, and robust operation in real-world scenarios, which are key characteristics of next-generation intelligent mobile robotic systems.

The review process for this thematic section was conducted by the Guest Editors. All submissions were subject to a rigorous double-blind peer-review process involving at least three independent reviewers. Manuscripts were evaluated on the basis of their technical merit, novelty, methodological rigor, quality of the state-of-the-art analysis, and overall contribution to the field of intelligent mobile robotics. Special consideration was given to studies presenting validated robotic systems and demonstrating their effectiveness in real-world deployment scenarios. Next, we provide a summary of the contributions included in this thematic section.

Osorio and Martinez-Carranza, present a vision-language navigation framework for Unmanned Aerial Vehicle (UAV) deployment

in an environment that is GPS-denied search-and-rescue. This proposed approach enables operators to issue free-form natural language instructions which are translated into image-grounded short-horizon waypoints instead of directly generated text actions that are combined with distance estimations to produce three-dimensional displacement commands, while adaptive step-length control and continuous replanning improve navigation under changing visibility, moving targets and emerging obstacles. This proposal experiments in high-fidelity simulated rescue scenarios, achieving an online XY RMSE of 0.098 meters and demonstrating shorter paths, reduced completion times, and reliable success rates compared with baseline methods. These results support the potential of vision-language models for obstacle-aware UAV navigation; nevertheless, real-world flight validations are necessary.

Martinez Vasquez et al. propose a two-stage motion planning methodology for car-like robots operating in environments with obstacles. In the proposal homotopy continuation is used to solve a nonlinear system formulated in the SE(2) configuration space incorporating vehicle position, orientation, and collision avoidance through a rigid body model and convex polygon barrier functions, resulting in a collision-free configuration sequence that is then transformed into a kinematically feasible trajectory using Reeds-Shepp curves and a randomized shortcutting procedure that reduces the total path length, in this sense comparisons with six sampling-based planners from the Open Motion Planning Library (OMPL) library across scenarios of increasing complexity show that the method achieves a 100% success rate and consistently produces shorter paths with competitive computation times in constrained environments. Its deterministic formulation also provides reproducible behavior, while validation

on physical platforms and extensions to dynamic obstacles remain future research directions.

Vázquez-Meza, et al. present a multimodal methodology for the real-time recognition of previously unseen destinations in drone delivery scenarios using only textual descriptions. This approach relies on generative models to generate candidate images of a target location in which are then organized through clustering to reduce the search space for online matching. In this sense, to compare these generated images with those captured by the drone's onboard camera, the method combines semantic visual embeddings from Contrastive Language-Image Pre-training (CLIP) with appearance-based features extracted by ResNet, yielding a robust similarity measure. Experimental results show a recognition accuracy of 0.80 with an average processing rate of 18 Hz, supporting the feasibility of the approach for real-time applications. Despite limitations related to image generation quality and local environmental representation, this study demonstrates the promise of combining generative and visual-recognition techniques for UAV navigation and delivery tasks.

Cabrera-Ponce, et al. propose a perspective-robust gate detection method for autonomous drone racing based on dual YOLO bounding-box annotations. Unlike conventional detectors, which poorly represent gate geometry under rotations with oblique viewpoints and scale variations, the proposed strategy recovers the four gate corners and reconstructs an inner polygon corresponding to the traversable region. This geometric representation improves detection accuracy without requiring additional segmentation modules. Quantitative and qualitative evaluations across three environments show that the method achieves a mean intersection over-union score of 0.95 while maintaining real-time performance of approximately 150 frames per second, meaning that outperform single bounding box and YOLO based segmentation alternatives. The results demonstrate its suitability for high-speed autonomous navigation, although its applicability is currently limited to square or rectangular structures that can be reconstructed from multiple bounding boxes.

Durán Contreras, et al. present an XGBoost-based system for automatically distinguishing epileptic from non-epileptic electroencephalogram (EEG) signals. Using 11,500 one-second segments from the Bonn Epileptic Seizure Recognition dataset, the proposed pipeline operates directly on raw temporal amplitudes and compensates for class imbalance without requiring manual feature engineering. The model achieved an AUC-ROC of 0.9953, an accuracy of 97.57%, and an F1-score of 0.9389 for the epileptic class, with stable performance across stratified cross-validation. Compared with Random Forest and Support Vector Machine models, XGBoost provided a comparable discriminative capacity, a more balanced sensitivity-specificity tradeoff, and substantially shorter training times. These results support its use as an efficient and interpretable screening tool for prioritizing clinically relevant EEG segments, particularly in resource-limited healthcare environments.

Navarrete-Rocha et al. explore an alternative way to classify driving styles. Instead of relying on external sensors or complex systems, the study exclusively uses data obtained from the vehicle's OBD-II port, seeking a balance between precision and low computational cost. To achieve this, a process was designed ranging from data cleaning and normalization, using Z-scoring, to the application of K-means as an unsupervised learning technique. The number of clusters ($k=3$) was determined using the elbow method, achieving a validation of 0.63 on the Silhouette Score from a set of 7,349 real observations. Results confirm that it is possible to successfully segment driver behaviour into three defined profiles (normal, moderate, and dynamic) using only internal engine parameters. In essence, this work provides a robust, non-intrusive, and highly scalable framework for driver monitoring, with clear potential for integration into advanced driver assistance systems and mobile robotics environments.

Méndez-López et al. propose DRGD (Drone Racing Gate Detector), a modular solution based on classic machine vision for locating doors in drone flight sequences, both indoors and outdoors. The work focuses on the design of a deterministic

system integrating standard components of color, morphology, geometry, and validation using the PnP (Perspective-n-Point) algorithm under a static configuration. The detector architecture operates on two levels: a first stage that transforms the frame information into proposals based on color and geometry, and a second stage that applies geometric constraints and PnP validation before determining the final bounding box. The evaluation was performed using a CPU protocol with strictly separate stages: first calibration, validation, parameter freezing, and finally a blind test on a proprietary dataset. The results show accurate performance in controlled indoor environments while maintaining real-time execution times. In the case of outdoor environments, although the system remains functional under conditions such as those in sequence SQ5, the drop in performance during sequence SQ7 underscores the dependence on the environment as a central challenge for future implementations in embedded systems. When compared to a YOLO11n-based line, the study provides a quantitative technical reference that positions DRGD as a tool that can be transparent and reproducible.

López et al. document the implementation and validation adapted to the bipedal robot Unitree G1 using a classic autonomous navigation scheme. The use of a biped presents additional challenges such as body oscillations, sensor instability, and odometry uncertainty; these factors complicate localization and trajectory tracking. An operating system is proposed under the ROS2 environment that integrates lidar-based perception, occupancy-grid mapping, AMCL localization, A*-based global planning, and velocity-based. The approach was evaluated in a controlled indoor environment with 15 tests along a 4.5-meter trajectory. The results show 100% success, no collisions, a mean time of 48.4 seconds, a positional error of 0.121 m, and an orientation error of 10.67° , confirming its viability.

Moreno et al. compare three object detection models for unrestricted landmarks in humanoid soccer player robots. The work evaluates the performance of YOLOv8, RT-DETR and

Deformable-DETR to provide accurate spatial observations in dynamic environments. The method uses the TORSO-21 dataset, augmented with images from the Robocup Humanoid League 2025, using the angular location error metric (in pixels) as an alternative to the IoU to evaluate the accuracy of the landmark center. The approach is evaluated using hardware inference tests with an NVIDIA GeForce RTX 3050 GPU, comparing accuracy, recall, angular error, and computation time. The results show that Deformable-DETR achieves the lowest angular localization error in all classes, although YOLOv8 offers the highest computational efficiency (83.82 FPS), highlighting that recall is the most critical metric for integration with filter localization systems.

Rojas-Perez et al. propose an unmanned aircraft system (UAS) to solve the status estimate in autonomous drone races (ADRs). The work introduces a lightweight, vision-only solution (monocular) that uses SLAM without requiring GPU acceleration. The method proposed uses monocular SLAM (ORB-SLAM2) with metric scale initialization through prior knowledge of door dimensions, along with a submap decomposition strategy to mitigate long-term drift and a no-scent Kalman filter (UKF) for state estimation. The approach is assessed by running autonomous flights in the controlled environment of the test arena of the second season of the A2RL championship, measuring track completion ability and track tracking accuracy. The results show that the system allowed the drone to complete the track autonomously at speeds of up to 3 m/s, highlighting that a monocular SLAM approach is an efficient and generalizable alternative to complex visual-inertial odometry or reinforcement learning solutions.

Guest Editors:

Juan Rubén Delgado Contreras
Carlos Eric Galván Tejada
Antonio García Domínguez
Juan Humberto Sossa Azuela