Instantaneous Dynamic-Obstacle Perception for Autonomous UAV Navigation

Md. Mukidur Rahman, Soo Young Shin*

mukidurrahman@kumoh.ac.kr, wdragon@kumoh.ac.kr*

Department of IT Convergence Engineering

Kumoh National Institute of Technology

Abstract

Navigating uncrewed aerial vehicles (UAVs) in human-co-located environments, such as construction sites and precision agriculture, requires perception systems capable of understanding complex dynamic behaviors. Conventional perception systems predominantly rely on historical kinematic data to predict future obstacle trajectories. However, these methods exhibit significant latency and fail to capture immediate state changes when facing sudden appearances or abrupt orthogonal turns of human obstacles. To address these limitations, this paper proposes a dynamic-obstacle perception framework that prioritizes instantaneous semantic cues over historical motion trends. We introduce an instantaneous semantic heading vector directly derives from a vision-based pose estimation network. This semantic heading vector allows the system to determine the direction of human movement and probable navigational intent, independent of historical displacement. Experimental results demonstrate that incorporating semantic pose attributes allows perception system to yield faster and more accurate heading estimation than purely kinematic baselines for downstream trajectory prediction and safety planning in cluttered dynamic environments.

Keywords: Human-heading, Human-trajectory, Dynamic obstacle avoidance, UAV navigation

Ⅰ. Introduction

Autonomous Unmanned Aerial Vehicles (UAVs) are increasingly utilized in human-co-located sectors, such as indoor construction and aerial cinematography [1]. Unlike navigation in static settings, these operations require robust perception systems capable of ensuring safety around moving agents like workers and machinery [5]. However, safe navigation in such cluttered environments remains a significant challenge due to the unpredictable nature of human movements and the limited computational resources available on lightweight aerial platforms [1].

Existing solutions primarily rely on onboard RGB-D sensors for dynamic obstacle detection and tracking [2], [3]. While recent approaches have improved detection through ensemble strategies or intent-based planning [4], [5], they predominantly depend on historical kinematic data to estimate future states. These methods often exhibit significant latency and struggle to react to abrupt directional changes, such as orthogonal turns. In these scenarios, historical kinematic data becomes insufficient or misleading.

To address these limitations, this paper presents a dynamic-obstacle perception framework that prioritizes instantaneous semantic cues over historical kinematic trends. Unlike traditional methods, we utilize vision-based pose estimation to derive a semantic heading vector directly from skeletal keypoints. This allows the system to determine navigational intent independent of physical displacement, enabling faster and more accurate trajectory prediction for safe UAV navigation in dynamic environments.

Ⅱ. Proposed System

As illustrated in Fig. 1, the proposed framework integrates perception, prediction, and planning for safe UAV navigation. The dynamic obstacle perception module ensembles geometric detectors (U-depth, DBSCAN) with a YOLO-Pose estimator to detect objects and extract the heading vector for immediate intent recognition. Detected agents are tracked and identified as dynamic obstacles and subsequently filtered from the map. Kinematic history fuses pose-derived semantic intent to forecast future states for generating collision-free trajectories.

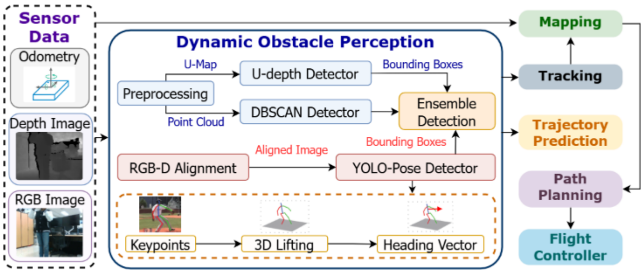

Fig.1: Dynamic-Obstacle Perception Framework

Ⅲ. Methodology

The perception pipeline utilizes a pretrained pose model to perform real-time top-down pose estimation. Our framework extracts four fiducial keypoints from the skeleton to define the torso plane: the left and right shoulders, and the left and right hips.

These raw 2D pixel coordinates $(u,\ v)$ are back-projected into 3D space utilizing aligned median filtered depth values Z_c. The corresponding 3D coordinate $P_c = [X_c,\ Y_c,\ Z_c]^T$ in the camera optical frame is then reconstructed using the pinhole camera model:

$$X_c = \frac{(u - c_x).Z_c}{f_x}, \qquad Y_c = \frac{(v - c_y).Z_c}{f_y}$$

where f_x, f_y represent the focal lengths and c_x, c_y denote the principal points of the camera intrinsic matrix.

The geometric human-heading is derived by constructing two orthogonal vectors relative to the

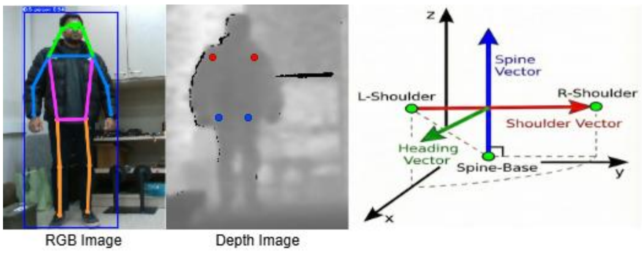

Fig.2: Illustration of Heading Estimation

torso. The horizontal shoulder vector V_{sh} between the shoulders, and vertical spine vector V_{sp} are calculated from 3D coordinates of the corresponding points. The raw heading vector N_{raw} is calculated as the cross product of the shoulder and spine vectors, resulting in a normal vector perpendicular to the torso plane.

$$\boldsymbol{N_{\mathrm{raw}}} = \boldsymbol{V_{\mathrm{sh}}} \times \boldsymbol{V_{\mathrm{sp}}}$$

To render this vector actionable, N_{raw} is transformed from the camera frame into the inertial world frame using the camera's extrinsic orientation matrix R_{cam} and projected onto 2D ground plane to yield the final heading H.

$$H = normalize\left(\begin{bmatrix} (R_{cam}.N_{raw})_x \\ (R_{cam} \cdot N_{raw})_y \\ 0 \end{bmatrix}\right)$$

To ensure trajectory stability against high-frequency pose jitter, we apply an Exponential Moving Average (EMA) filter. This includes a consistency check where the dot product between the previous and current heading is evaluated. The filter resets for sudden turn or pose flip when the directional change exceeds 90° otherwise, the standard smoothing is applied:

$$\boldsymbol{H_t = \alpha \cdot H_{new} + (1-\alpha).H_{t-1}}$$

Finally, the yaw angle required for trajectory prediction is derived via the arctangent of the heading components.

IV. Results

The proposed method was implemented on ROS platform using an Intel RealSense D435 camera and yolo11n-pose. RViz visualizations in Fig. 3 confirm the system performance. The ensemble detector successfully identifies dynamic obstacles within blue bounding boxes. Furthermore, the system accurately computes the heading vector for various human orientations. These vectors appear as blue arrows and demonstrate the accurate estimation of directional intent. The runtime of the system modules is detailed in Table I, with measurements conducted on Intel NUC.

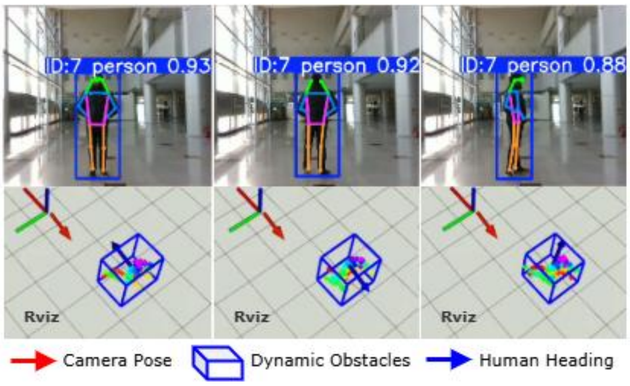

Fig.3: Dynamic Detection with Human-heading

TABLE I: RUNTIME OF MODULES OF THE SYSTEM

System Module	Runtime (ms)
Heading estimation	0.015
YOLO-Pose detection	15.23

V. Conclusion and Future Work

This paper presents an instantaneous dynamic - obstacle perception framework for autonomous UAV navigation. We utilized skeletal keypoints to derive instantaneous heading vectors. This geometric approach addresses the latency inherent in kinematic-based intent estimation. Real-world experiments demonstrated that our method yields fast and accurate directional cues of human movements. Future research will focus on integrating this semantic heading with human-trajectory prediction model to enhance autonomous UAV navigation in complex dynamic environments.

ACKNOWLEDGEMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2024-00437190, 50%) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2025-RS-2022-00156394, 50%) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

REFERENCES

[1] Liang, H., S. C. Lee, W. Bae, J. Kim, and S. Seo. Towards UAVs in Construction: Advancements, Challenges, and Future Directions for Monitoring and Inspection. Drones, 7 (3), 202. 2023.

[2] Lu, Minghao, Han Chen, and Peng Lu. "Perception and avoidance of multiple small fast moving objects for quadrotors with only low-cost RGBD camera." IEEE Robotics and Automation Letters 7, no. 4 (2022): 11657-11664.

[3] Lin, Jiahao, Hai Zhu, and Javier Alonso-Mora. "Robust vision-based obstacle avoidance for micro aerial vehicles in dynamic environments." In 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 2682-2688. IEEE, 2020.

[4] Xu, Zhefan, Xiaoyang Zhan, Yumeng Xiu, Christopher Suzuki, and Kenji Shimada. "Onboard dynamic-object detection and tracking for autonomous robot navigation with rgb-d camera." IEEE Robotics and Automation Letters 9, no. 1 (2023): 651-658.

[5] Xu, Zhefan, Hanyu Jin, Xinming Han, Haoyu Shen, and Kenji Shimada. "Intent prediction-driven model predictive control for uav planning and navigation in dynamic environments." IEEE Robotics and Automation Letters (2025).