RoboTC: a robotic platform for privacy-ware active thermal comfort data collection

Weijia Cai[1] and Zhengbo Zou[2]

***Abstract*— Thermal comfort commissioning requires pairing objective environmental measurements with subjective occupant feedback collected at the same location. However, obtaining such paired data using human inspectors is labor-intensive and difficult to scale. We present RoboTC, a robotic system that actively collects subjective feedback on occupants' thermal preferences. RoboTC comprises four modules: Human Detector, Human Localization, Human-Aware Approach Planner, and Interaction Interface. To train the human detector, we collect and annotate a multimodal dataset consisting of 10K paired depth and thermal images covering indoor activities defined by ASHRAE. We conduct experiments to evaluate both human detection and human approach performance. The human detector achieves an IoU of 0.75, a precision of 91.49%, and a recall of 80.13%. To evaluate the human-aware approach planner, we conduct real-world experiments across ten different locations, achieving a 90% success rate.**

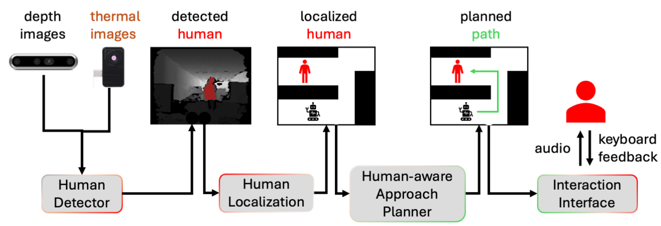

Fig. 1. RoboTC system overview.

I. INTRODUCTION

Buildings account for approximately 33% of global energy consumption [1], while Heating, Ventilation, and Air Conditioning (HVAC) systems make up 38% of building energy consumption [2] to maintain indoor thermal comfort. Traditionally, HVAC control is based on setpoint schedules [3]. It is common practice for facility managers to conservatively assume full occupancy during operating hours, causing unreasonably high energy cost and low thermal comfort. This is mainly due to the lack of real-time information about occupancy and thermal preference of the occupants [4]. Recent studies show that collecting objective environment quality measurements (e.g., temperature) as well as subjective thermal preference from occupants in different locations, can save 16%-32% total building energy [5].

In practice, however, collecting both the objective and subjective measurements with accurate spatial information is challenging and inefficient for human inspectors [4]. Recent advances in automating Indoor Environment Quality (IEQ) data collection using robots show a promising path to close the missing link between the HVAC control and in-time indoor environmental conditions [6], [7], [8], [9]. For instance, Jin et al. [6] proposed a framework that automates IEQ monitoring using robots, showing that the robotic platform can provide accurate and high-resolution measurements. However, the robot traverse route was predefined, which is less versatile for dynamic indoor environments. Cai et al. [9] proposed a robotic system for efficient data collection and spatiotemporal maps of air pollutants. However, they do not provide active collection of human preference for thermal comfort.

In this paper, we propose **RoboTC**, a robotic system that automates thermal comfort commissioning for both objective environmental measurements and human thermal preference. **RoboTC** is composed of four modules, including Human Detector, Human Localization, Human-aware Approach Planner, and Interaction Interface. Our main contributions are:

- We develop a robot platform that provides active human thermal preference reporting.
- We collect and annotate a multimodal dataset including 12K depth and registered thermal image pairs for human detection training and testing.
- We trained a privacy-aware human detection model based on depth and thermal images only.

II. RoboTC

In this section, we introduce **RoboTC**, a robotic system that actively collects occupants' thermal preferences together with localized environmental measurements, shown in 1.

A. Hardware System Design

We develop a robot platform with multiple sensors for localization and human detection, a mobile base for navigation, a laptop for computing and audio output, and a 3-key keyboard for receiving human feedback. Shown in Fig. 2, a Realsense D455 depth camera and a thermal imaging camera are used to detect human, while a Unitree L2 LiDAR is used to perform Simultaneous Localization and Mapping (SLAM) for robot localization and environment mapping. A 3-key keyboard is used to receive human feedback about their thermal preference. We adopt a mobile base design from LeKiWi [10] powered by a power station. A laptop provides computing resources for mobile base commanding,

*This work was not supported by any organization

[1]Weijia Cai is with the Department of Civil Engineering and Engineering Mechanics, Columbia University, West 120th Street, New York, NY, USA wc2900@columbia.edu

[2]Zhengbo Zou is with Faculty of Civil Engineering and Engineering Mechanics, Columbia University, West 120th Street, New York, NY, USA zz3279@columbia.edu

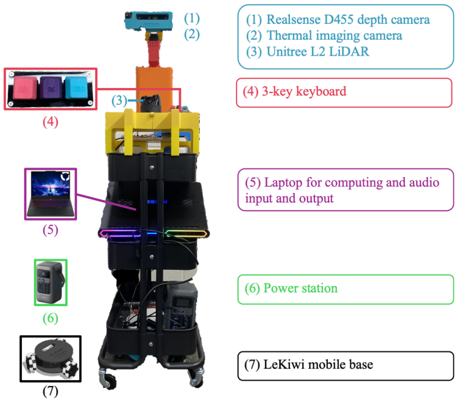

Fig. 2. Robot Hardware.

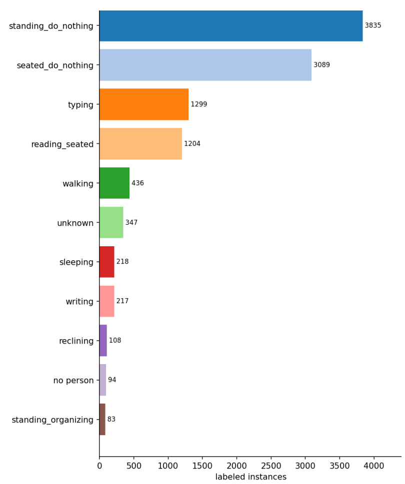

Fig. 3. Human activity distribution in multimodal Dataset

real-time human detection and approaching, and audio output and keyboard feedback for human robot interaction.

B. Human Detector

To avoid privacy issues, we develop a human detector based on depth and thermal images, which do not preserve recognizable facial information from humans. We first collect and annotate a multimodal dataset including depth and thermal image pairs with the segmentation mask of human. Then we design an ensemble model trained on this dataset for human detection.

1) Multimodal Dataset: To collect a comprehensive dataset, we collect different indoor human activities based on typical office tasks from ASHRAE standard 55 [11], in 14 different locations, with 9 different participants, resulting in 10945 pairs of depth and thermal images, as shown in Fig. 3. For the annotation process, we utilize a pretrained Visual Language Model [12] and SAM3 [13] to obtain the human activity labels and the segmentation labels. We then manually verify the annotation results. We calibrated the intrinsic and extrinsic camera parameters of both cameras so that segmentation masks from the depth camera frame can be projected to the thermal camera frame.

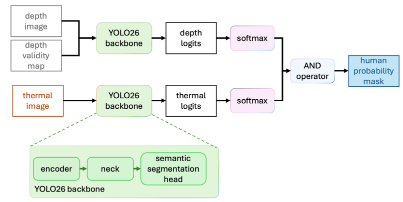

Fig. 4. Ensemble Model Architecture for Human Detection

2) Human Detector Model: We develop an ensemble model that takes depth images, depth validity maps, and thermal images as inputs, as shown in Fig. 4. The depth branch receives a depth image stacked with its corresponding validity map, which indicates valid depth pixels, while the thermal branch receives the registered thermal image. Each branch is processed by an independent YOLO26 backbone [14], whose dense logits are converted into per-pixel class probabilities using a softmax layer. We adopt a conservative ensemble strategy to suppress false-positive detections. Specifically, the probability maps from the two branches are thresholded and combined using a logical AND operation, retaining only pixels classified as human by both modalities to produce the final human mask. This conjunctive fusion is deliberately conservative because the depth and thermal modalities tend to fail under different conditions; requiring agreement between them therefore suppresses modality-specific false positives before they propagate to downstream localization and planning. Reducing false positives is particularly important because erroneous human detections could cause the robot to execute unnecessary or unsafe approach maneuvers, potentially increasing the risk of collision.

C. Human Localization

To localize a human in a 3D map, the Human Localization module consists of three steps: 1) SLAM, 2) depth-to-point-cloud transformation, and 3) human position registration. For SLAM, we implement a LiDAR-based method [15] that takes LiDAR measurements as input and estimates the robot's position and orientation in real time. For the depth-to-point-cloud transformation, we convert each depth image into a local point cloud following the method in [16]. Finally, we crop the local point cloud using the human mask generated by the Human Detection module and transform the resulting human point cloud into the global coordinate frame using the robot pose estimated by SLAM.

D. Human-aware Approach Planner

Algorithm 1 describes the safety-aware human approach behavior. When a human is detected, the robot first calculates an approach goal pose and then plans a path from its current pose to the goal, prioritizing safety over path length. We deploy a dynamic window approach (DWA) [17] as local path planner to generate a safe path for the robot execute. For the path execution, we utilize Lekiwi controller [10] based on kinematics, shown as $\textsc{GoTo}(n)$. We utilize the occupancy map $\mathcal{M}$, 2D probability map of occupied objects, to enable early stop for potential collision.

Algorithm 1 Human-aware Approach Planner

$\mathcal{M}$: occupancy map.
g: robot navigational goal.
q_{cur}: current robot pose.
$\mathcal{P}$: planned path consisting of a list of waypoints.
$occu_{max}$: occupancy threshold;
s: task status.
if a human h is detected **then**
    $g \leftarrow$ HUMANLOCALIZATION$(h, \mathcal{M})$ ▷ II-C
    $\mathcal{P} \leftarrow$ PLANPATH$(q_{\text{cur}}, g, \mathcal{M})$ ▷ [17]
    $s \leftarrow$ EXECUTEPATH$(\mathcal{P})$
end if
procedure EXECUTEPATH($\mathcal{P}$, $\mathcal{M}$, q_{cur})
    for each node $n \in \mathcal{P}$ **do**
        GOTO(n)
        if $\mathcal{M}(q_{curr}) > occu_{max}$ **then**
            return COLLISION
        end if
    end for
    return SUCCESS
end procedure

E. Interaction Interface

To enable efficient interaction with occupants, we deploy an Interaction Interface module that plays an audio prompt when the robot approaches an occupant and collects feedback through a three-key keypad with three options: "HOT"," OK", and "COLD". Specifically, **RoboTC** plays the audio prompt: "Hi, how do you feel about the temperature here?" and waits for the occupant to select a response. This interface enables rapid and intuitive collection of occupants' thermal preferences with minimal interaction effort.

III. EXPERIMENTS AND RESULTS

A. Experiment Design

The goal of this paper is to develop a safe and robust robotic system for detecting and approaching humans to collect thermal comfort data. We design separate experiments to evaluate the human detection and approaching components. For human detection, we collect and annotate a test dataset of 1,843 frames spanning five unseen locations. The dataset is deliberately constructed to capture failure modes that motivate our two-modality design, including hot objects with temperatures close to that of human skin, thermal reflections of people on specular surfaces, and photographs of people mounted on walls. We compare the ensemble model against three baselines. Two are single-modality models that use either thermal or depth images as input, while the third is an early-fusion model that combines depth and thermal features through a gated fusion layer [18] before passing them to a single backbone. All baselines use the same YOLO26 backbone as an individual branch of the ensemble model and differ only in their input modalities and fusion strategies. For human approaching evaluation, we design ten test cases in ten different locations, covering a range of human activities. In each test case, the robot is required to approach a human and reach a predefined safe distance. An approach is considered successful if the robot reaches this distance without collision.

TABLE I
HUMAN DETECTION RESULTS.

	Training			*Testing*		
Model	IoU	Prec.	Rec.	IoU	Prec.	Rec.
Ensemble	0.9400	**0.9766**	0.9679	**0.7457**	0.9149	0.8013
Depth	**0.9452**	0.9723	**0.9791**	0.7394	**0.9277**	0.7846
Thermal	0.9229	0.9602	0.9646	0.6421	0.7741	0.7902
Early-fusion	0.9412	0.9688	0.9759	0.7271	0.8770	**0.8097**

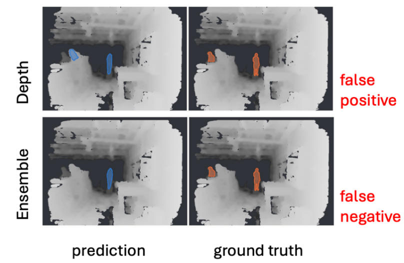

Fig. 5. Qualitative Comparison between Ensemble and Depth only Models.

B. Human Detection Results

We show the model performance on the training and test dataset in Table I. On the training set, all models achieve relatively high performance, with IoU values above 0.92 and precision and recall above 0.96. The depth-only model achieves the highest training IoU (0.9452) and recall (0.9791), while the ensemble model achieves the highest precision (0.9766). On the test set, performance decreases for all models, reflecting the increased difficulty of generalizing to unseen locations and challenging scenarios. The ensemble model achieves the highest test IoU of 0.7457, outperforming other models. Although the depth-only model achieves slightly higher precision than the ensemble model on the test set, the qualitative results in Fig. 5 show that it remains susceptible to modality-specific false-positive detections under challenging conditions. Such false positives are particularly undesirable in our application, as an incorrectly detected human may trigger an unnecessary approaching maneuver and increase the risk of unsafe robot behavior; in contrast, the ensemble model suppresses this false positive by requiring agreement between the depth and thermal modalities.

C. Human Approaching Results

Fig. 6 shows two representative outcomes. In the top row, the robot correctly detects a seated human and comes to a stop within the safe interaction range. The bottom row shows a failure case in which detection succeeds but the robot stops short of the target range after the collision check is triggered by a nearby wall. Across all trials, the robot achieves a 90% approach success rate, with an average traveled distance of 4.78 m and an average travel duration of

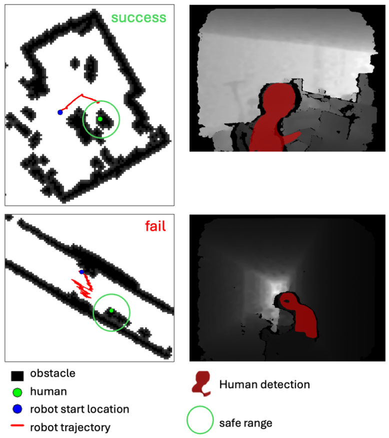

Fig. 6. Qualitative Results for Human Approaching.

3.23 min. The observed failure results from the conservative collision-avoidance mechanism rather than a detection or path-planning error.

IV. CONCLUSION AND FUTURE WORKS

In this paper, we propose **RoboTC**, a robotic system that automates thermal comfort commissioning by collecting subjective occupants' thermal preferences. Our contributions are threefold. First, we design a robotic platform that actively collects subjective thermal comfort data. Second, we release an open-source multimodal dataset collected under typical indoor activities specified in ASHRAE standards. Third, we design and implement a privacy-aware human detection and approach method that enables the robot to safely approach occupants and collect their subjective thermal preferences. Future work will focus on expanding the multimodal dataset, conduct more comprehensive human-robot interaction tests, and developing a more socially aware human approach strategy.

REFERENCES

[1] IEA, "Tracking clean energy progress 2023," International Energy Agency, Paris, France, Tech. Rep., 2023, licence: CC BY 4.0. [Online]. Available: https://www.iea.org/reports/tracking-clean-energy-progress-2023
[2] M. González-Torres, L. Pérez-Lombard, J. F. Coronel, I. R. Maestre, and D. Yan, "A review on buildings energy information: Trends, end-uses, fuels and drivers," *Energy Reports*, vol. 8, pp. 626–637, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484721014 27X
[3] M. Fadzli Haniff, H. Selamat, R. Yusof, S. Buyamin, and F. Sham Ismail, "Review of hvac scheduling techniques for buildings towards energy-efficient and cost-effective operations," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 94–103, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032113004322
[4] Z. Nagy, B. Gunay, C. Miller, J. Hahn, M. M. Ouf, S. Lee, B. W. Hobson, T. Abuimara, K. Bandurski, M. André, C.-L. Lorenz, S. Crosby, B. Dong, Z. Jiang, Y. Peng, M. Favero, J. Y. Park, K. Nweye, P. Nojedehi, H. Stopps, L. Sarran, C. Brackley, K. Bassett, K. Govertsen, N. Koczorek, O. Abele, E. Casavant, M. Kane, Z. O'Neill, T. Yang, J. Day, B. Huchuk, R. T. Hellwig, and M. Vellei, "Ten questions concerning occupant-centric control and operations," *Building and Environment*, vol. 242, p. 110518, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360132323005450
[5] E. Mills, "Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the united states," *Energy Efficiency*, vol. 4, no. 2, pp. 145–173, May 2011.
[6] M. Jin, S. Liu, S. Schiavon, and C. Spanos, "Automated mobile sensing: Towards high-granularity agile indoor environmental quality monitoring," *Building and Environment*, vol. 127, pp. 268–276, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360132317305012
[7] W. Cai, L. Huang, and Z. Zou, "Leveraging open-vocabulary object detection in goal-oriented navigation for assessing indoor energy–intensive devices," *Journal of Computing in Civil Engineering*, vol. 39, no. 5, p. 04025060, 2025. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/JCCEE5.CPENG-6376
[8] W. Cai, L. Huang, and Z. Zou, "Roboauditor-llm: A robotic system for energy auditing via a large language model and an actionable semantic feature map," *ASCE OPEN: Multidisciplinary Journal of Civil Engineering*, vol. 3, no. 1, p. 04025013, 2025. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/AOMJAH.AOENG-0087
[9] W. Cai and Z. Zou, "Usiva: Uncertainty-guided spatiotemporal mapping for indoor ventilation assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2026, pp. 9440–9449. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2026W/CV4AEC/html/Cai_USIVA_Uncertainty-guided_Spatiotemporal_Mapping_for_Indoor_Ventilation_Assessment_CVPRW_2026_paper.html
[10] P. Kooijmans, M. Chandaka, B. Chandaka, G. Wang, and A. Patel, "LeKiwi: A low-cost mobile manipulator," https://github.com/SIGRobotics-UIUC/LeKiwi, 2025, software. Accessed: 2026-08-15.
[11] ASHRAE, *ANSI/ASHRAE Standard 55-2023: Thermal Environmental Conditions for Human Occupancy*, American Society of Heating, Refrigerating and Air-Conditioning Engineers Std., 2023, supersedes ANSI/ASHRAE Standard 55-2020. [Online]. Available: https://www.ashrae.org/technical-resources/bookstore/standard-55-thermal-environmental-conditions-for-human-occupancy
[12] Qwen Team, "Qwen3-VL-4B-Instruct," https://huggingface.co/Qwen/Qwen3-VL-4B-Instruct, 2025, model weights, Hugging Face. Accessed: 2026-08-15.
[13] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris Coll-Vinent, C. Ryali, K. V. Alwala, H. Khedr, A. Huang, J. Lei, T. Ma, B. Guo, A. Kalla, M. Marks, J. Greer, M. Wang, P. Sun, R. Rädle, T. Afouras, E. Mavroudi, K. Xu, T.-H. Wu, Y. Zhou, L. Momeni, R. HAZRA, S. Ding, S. Vaze, F. Porcher, F. Li, S. Li, A. Kamath, H. K. Cheng, P. Dollar, N. Ravi, K. Saenko, P. Zhang, and C. Feichtenhofer, "Sam 3: Segment anything with concepts," in *International Conference on Learning Representations*, C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust, Eds., vol. 2026, 2026, pp. 138 846–138 923. [Online]. Available: https://proceedings.iclr.cc/paper_files/paper/2026/file/e0982cbc81401df3430ee1ff780dc7a2-Paper-Conference.pdf
[14] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, "Ultralytics YOLO26: Unified real-time end-to-end vision models," 2026. [Online]. Available: https://arxiv.org/abs/2606.03748
[15] D. He, W. Xu, N. Chen, F. Kong, C. Yuan, and F. Zhang, "Point-lio: Robust high-bandwidth light detection and ranging inertial odometry," *Advanced Intelligent Systems*, vol. 5, no. 7, p. 2200459, 2023. [Online]. Available: https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202200459
[16] W. Cai, L. Huang, and Z. Zou, "Actively-exploring thermography-enabled autonomous robotic system for detecting and registering hvac thermal leaks," *Automation in Construction*, vol. 152, p. 104901, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580523001619
[17] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The office marathon: Robust navigation in an indoor office environment," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 300–307.
[18] Y. Cheng, R. Cai, Z. Li, X. Zhao, and K. Huang, "Locality-sensitive deconvolution networks with gated fusion for rgb-d indoor semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1475–1483.