Distributional Semantics for Robust Global Localization in Cluttered, Geometrically Aliased Environments

Shivendra Agrawal, Alessandro Roncone, and Bradley Hayes

***Abstract*— Many indoor workspaces such as warehouses, laboratories, and retail spaces are *quasi-static*: their global geometric layout remains permanent, but local semantics change continually. These spaces often have repetitive geometry, dynamic clutter, and perceptual noise that makes standard vision-based localization brittle. We present Distributional Semantic Monte Carlo Localization (DS-MCL), a particle filter for robust global localization that treats scene semantics as *statistical evidence over object categories* rather than fixed-quantity landmarks. DS-MCL fuses a geometric depth likelihood with a category-centric semantic similarity, utilizing a precomputed bank of semantic viewpoints to perform *inverse semantic proposals* for fast, targeted hypothesis generation on low-cost hardware. We evaluate DS-MCL across two environments. In a rigorously controlled quasi-static environment, DS-MCL achieves a 97% global localization success rate, heavily outperforming geometric and fixed-semantic baselines. Furthermore, in a 3,500 sq. ft. operational retail store, leveraging an open-vocabulary vision pipeline, DS-MCL significantly outperforms fixed-quantity baselines (62% vs 42% success). By modeling semantics distributionally, DS-MCL resolves geometric aliasing and provides an infrastructure-free building block for reliable autonomy in dynamic real-world environments.**

I. INTRODUCTION

Many real-world indoor environments are *quasi-static*: their global geometric layout (e.g., walls, structural shelving) remains permanent over long time horizons, but their local semantic contents change continually. In these settings, traditional geometry-based particle filters like Adaptive Monte Carlo Localization (AMCL) [1] suffer from severe perceptual aliasing due to visually repetitive geometry. Furthermore, dynamic clutter and noisy observations routinely violate static map assumptions, leading to rapid particle impoverishment [2]. Consequently, robust, start-anytime global localization in these GPS-denied environments remains a critical open problem for reliable robot autonomy.

To address these challenges, we present Distributional Semantic Monte Carlo Localization (DS-MCL, Fig. 1), a semantic particle filter tailored for robust global localization. Unlike prior explicit semantic approaches that treat detected objects as discrete landmarks with fixed quantities [3]–[6], DS-MCL models semantic features as probabilistic distributions over object counts and spatial arrangements. This representation captures the intrinsic variability of real-world environments, preserving the statistical structure necessary for stable observation-driven localization even as local object configurations continuously fluctuate.

The authors are with the Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309, USA. {shivendra.agrawal, alessandro.roncone, bradley.hayes}@colorado.edu

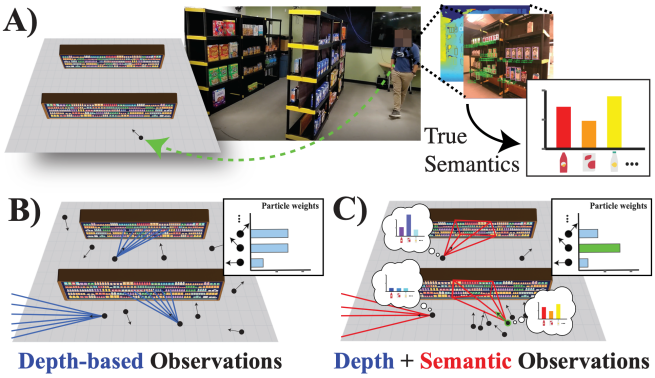

Fig. 1. An overview of DS-MCL in a quasi-static environment. Traditional depth-based observation models struggle with geometric aliasing in repetitive aisles (B). DS-MCL introduces an inverse semantic model that injects targeted particles based on probabilistic semantic cues (C), breaking the aliasing and enabling robust global localization.

Our core technical contributions are twofold:

- A semantic mapping and observation design that encodes object collections as statistical distributions over counts and arrangement, providing inherent robustness to semantic perturbations and geometric ambiguity.
- A real-time, inverse observation model-based particle filter that leverages this representation on low-cost, vision-based hardware to efficiently propose and evaluate high-quality global pose hypotheses.

II. RELATED WORK

A. Vision and Semantics-Based Localization

Particle filter-based localization methods such as Monte Carlo Localization (MCL) and its adaptive variant AMCL [1] remain standard for onboard robot localization due to their scalability and integration with modern navigation frameworks [7]. However, their reliance on static geometric maps and depth data makes them brittle in quasi-static environments where geometry is repetitive or transient.

This limitation has driven research toward semantic localization [8]. Earlier work augments geometric maps with explicit, object- or region-level labels that serve as long-term landmarks for place recognition and drift reduction [4]–[6]. Text-based cues have also proven effective as distinctive features in structured indoor spaces [3]. More recent efforts adopt implicit semantic representations, encoding spatial and semantic structure jointly in learned neural fields [9]. However, these methods often assume stable semantic identities

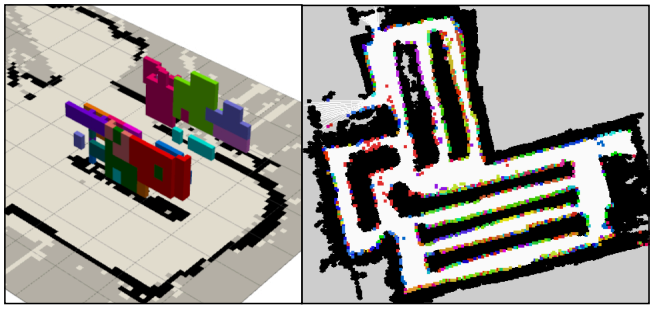

Fig. 2. The semantic map overlaid on the occupancy grid. Each voxel stores a distribution over object class counts. Ray casting on this layer yields the expected semantic vector comprising class counts, distances, and angles. *Left:* Controlled Environment, *Right:* Operational Retail Store.

and fixed counts, assumptions that rarely hold in quasi-static domains where objects are moved, occluded, or replaced.

Recent deep visual localization methods [10], [11] demonstrate impressive accuracy but require server-grade compute, precluding real-time use on wearable edge devices. Furthermore, while modern visual SLAM [12] excels at local tracking, it typically requires extensive exploration to achieve global localization against a prior map, limiting instantaneous recovery. In contrast, DS-MCL treats environment contents as statistical, count-based semantic distributions, directly addressing quasi-static dynamics while remaining lightweight enough for compute-constrained platforms.

III. OUR APPROACH

DS-MCL builds on the particle filter framework, augmenting standard depth likelihoods with a probabilistic semantic observation model. The joint observation model factors geometric and semantic likelihoods as:

$$p(\mathbf{z}_t \mid \mathbf{x}_t, m) = \eta \; p_d(\mathbf{z}_t^d \mid \mathbf{x}_t, m_d) \; p_s(\mathbf{z}_t^s \mid \mathbf{x}_t, m_s), \qquad (1)$$

where η is a normalizer, m_d and m_s are the geometric and semantic maps respectively, and $\mathbf{z}_t = (\mathbf{z}_t^d, \mathbf{z}_t^s)$ are the depth and semantic observations.

A. Semantic Mapping and Distributional Representation

DS-MCL maintains a hybrid map combining a 2D occupancy grid of traversable space built using off-the-shelf mapping algorithms [13] with a coarse 3D semantic layer (Fig. 2). The semantic map discretizes the volume into 20×20 cm cells in (x, y) and 30 cm in z, a resolution selected to balance real-time ray-casting efficiency with the typical spatial footprint of retail inventory. Rather than storing discrete object centroids, each voxel stores a running distribution over observed object *classes* and their *counts*.

To demonstrate perception-agnostic scalability, we populate this map using two distinct pipelines:

a) Pipeline A (Closed-Set Semantics): For controlled benchmarking with a known number of classes, we utilize a YOLOv9 detector fine-tuned on the SKU-110K dataset [14] to extract generic bounding boxes. A ResNet50-based classifier then assigns these to manually defined application-level classes. Grouping underlying item variations into a small category set provides robustness against local churn.

b) Pipeline B (Self-Supervised Category Discovery): To demonstrate zero-shot scalability for real-world deployment, we combine YOLOv9's class-agnostic proposals with a pre-trained DINOv3 Vision Transformer to automatically discover latent semantic categories via clustering, bypassing the need for manual annotation.

Objects detected via either pipeline are projected into the map frame. Given pixel coordinates $\tilde{\boldsymbol{u}} = [u, v, 1]^\top$ at the bounding box's median depth Z_c, the 3D world position $\boldsymbol{X}_w$ is computed as $\boldsymbol{X}_w = \boldsymbol{t}_{wc} + Z_c \boldsymbol{R}_{wc} \boldsymbol{K}^{-1} \tilde{\boldsymbol{u}}$, where $\boldsymbol{K}$ is the camera intrinsic matrix, and $\boldsymbol{R}_{wc}$ and $\boldsymbol{t}_{wc}$ denote the camera-to-world rotation matrix and translation vector, respectively. We refine each estimated position along the camera ray using Bresenham's line algorithm [15] until it aligns with the nearest occupied cell in the geometric map, preventing minor misalignments due to depth noise.

B. Semantic Observation Model

DS-MCL generates a semantic observation $\mathbf{z}_t^s$ from the live camera view. This concatenates three sub-vectors: (i) a class-count vector $\mathbf{v}_c$, (ii) a mean range vector $\mathbf{v}_d$, and (iii) a mean bearing vector $\mathbf{v}_\theta$.

To compare live and expected observations, we define a composite similarity:

$$S(\mathbf{z}^s, \hat{\mathbf{z}}^s) = \alpha S_{\text{counts}} + \beta S_{\text{distance}} + \gamma S_{\text{angle}} \qquad (2)$$

We normalize $\mathbf{v}_c$ to form a categorical distribution and compare it via Jensen–Shannon divergence (JSD). Defining P and Q as the normalized count distributions and $M = \frac{1}{2}(P+Q)$, we set $S_{\text{counts}} = 1 - \text{JSD}$ with:

$$\text{JSD}(P \parallel Q) = \sqrt{\frac{1}{2} \sum_{i=1}^{C} \left(P(i) \log \frac{P(i)}{M(i)} + Q(i) \log \frac{Q(i)}{M(i)} \right)}. \qquad (3)$$

By comparing normalized distributions rather than absolute counts, this formulation provides inherent robustness to perception failures. The spatial components $\mathbf{v}_d$ and $\mathbf{v}_\theta$ are compared using L2 distances. We map the unbounded range error d_{distance} to $S_{\text{distance}} = 1/(1 + d_{\text{distance}})$, while the bounded angular error d_{angle} is scaled by the camera's field-of-view (FOV) as $S_{\text{angle}} = 1 - (d_{\text{angle}}/\text{FOV})$.

C. Depth Observation Model

We use the depth image to incorporate geometric information. Given the expected range $\hat{z}^{(k)}$ for the k-th beam computed using CDDT-accelerated ray casting [16], we use the beam end-point mixture [1]:

$$p_d(z_t^{d,(k)}) = w_{\text{h}} \, p_{\text{hit}} + w_{\text{s}} \, p_{\text{short}} + w_{\text{m}} \, p_{\text{max}} + w_{\text{r}} \, p_{\text{rand}}, \qquad (4)$$

incorporating local Gaussian noise (p_{hit}), dynamic obstacles (p_{short}), and sensor failures ($p_{\text{max}}, p_{\text{rand}}$).

D. Inverse Semantic Model for Global Localization

The key innovation of DS-MCL is an *inverse semantic model* that proposes high-quality pose hypotheses directly from live semantic observations. We precompute expected semantic observations $\hat{\mathbf{z}}^s(\boldsymbol{x}, m_s)$ for discretized poses over the free space and cache them in a hashmap. We also build a reverse index mapping each product class to the poses from which it is visible. Algorithm 1 summarizes the online execution of the particle filter.

Algorithm 1 Online Localization with DS-MCL

Input:
Current sensor data: image i_t, depth d_t, odometry o_t
Pre-computed map data: semantic vectors $\hat{\mathbf{z}}^s(\mathbf{x})$
Output: Updated particle set P_{t+1}
Initialize: Particle set P_t sampled uniformly over free space

while True **do**
  $P_t \leftarrow \text{MotionModel}(P_t, o_t)$
  $z_t^s \leftarrow \text{GenerateSemanticVector}(i_t, d_t)$
  $\hat{x} \leftarrow \text{ExpectedPose}(P_t, W_t)$
  $\hat{z}_t^s \leftarrow \text{CalculateExpectedSemanticObs}(\hat{x})$
  $\text{sim} \leftarrow \text{CalculateSemanticSimilarity}(z_t^s, \hat{z}_t^s)$
  if $\text{sim} < \tau_{\text{sim}}$ **and** $\|z_t^{s,\text{count}}\|_1 > \tau_\kappa$ **then**
    $C \leftarrow \text{GetObservedClasses}(\mathbf{z}_t^s)$
    $\text{Candidates} \leftarrow \bigcup_{c \in C} \texttt{class\_to\_poses}[c]$
    $\text{Scores} \leftarrow [S(\mathbf{z}_t^s, \hat{\mathbf{z}}^s(\mathbf{x})) \text{ for } \mathbf{x} \in \text{Candidates}]$
    $\text{TopPoses} \leftarrow \text{GetTopK}(\text{Candidates}, \text{Scores}, k)$
    $P_t \leftarrow \text{InjectParticles}(P_t, \text{TopPoses})$
    $W_{\text{depth}} \leftarrow \text{DepthObservationModel}(P_t, d_t)$
    $W_{\text{sem}} \leftarrow \text{SemanticObservationModel}(P_t, i_t, d_t)$
    $W_t \leftarrow W_{\text{depth}} \odot W_{\text{sem}}$
  else
    $W_t \leftarrow \text{DepthObservationModel}(P_t, d_t)$
  end if
  $W_t \leftarrow \text{NormalizeWeights}(W_t)$
  $P_t \leftarrow \text{Resample}(P_t)$
  $P_{t+1} \leftarrow P_t$
end while

IV. HARDWARE AND EXPERIMENTAL EVALUATION

A. Hardware and Form Factors

To meet the constraints of low-profile platforms, DS-MCL uses compact vision sensors: an Intel RealSense D455 RGB-D camera and a T265 VIO camera for odometry, evaluated across wearable (chest-mounted) and cart-mounted configurations. We adopt these form factors to reflect practical compute constraints and to stress-test the method under depth-camera noise (lower scan frequency and point density than LiDAR) [17]. Importantly, because wearable and compact cart setups preclude the use of reliable wheel encoders, the system relies entirely on the T265's Visual-Inertial Odometry for particle propagation, making it a purely vision-based localization stack. Onboard processing was conducted on a Dell G15 laptop (Intel Core i7-11800H, NVIDIA RTX 3060). The full localization pipeline achieves real-time throughput of 9.6Hz without requiring LiDAR or wheel encoders. For the wearable use case, we intentionally avoid the usability barriers of handheld devices [18] and the social stigma of head-mounted systems [19].

B. Baselines and Metrics

We benchmark against two purely geometric baselines: Monte Carlo Localization (MCL) and its adaptive variant (AMCL) [1]. We also compare against a Fixed Quantity Landmark (FQL) baseline, which represents standard semantic localization approaches [3], [4] by treating detected object classes as having fixed, deterministic counts. For fairness, all methods utilized 1,500 particles and the same odometry and RGB-D streams.

Metrics and Convergence Criteria. Following established conventions, we evaluate DS-MCL via:

- **Global Localization Success:** A trial is a success if convergence occurs within the first 95% of the trajectory and remains converged until the end.
- **Tracking Success:** The percentage of trials that converge and remain strictly within the convergence thresholds until the end of the sequence.
- **Convergence Time (s):** The duration from initialization to the beginning of the final convergence for successful trials.
- **Global RMSE (G-RMSE):** The Absolute Trajectory Error (translation and rotation) computed post-convergence for trials that successfully globally localized.

We declare convergence when the estimated pose is within 0.7 m translation (representing a standard human standing reach [20]) and $\pi/4$ rad rotation of the ground truth.

C. Domain A: Controlled Cluttered Environment

To validate our approach against absolute ground truth, we first evaluate in an environment instrumented with an OptiTrack motion-capture system. The space was densely stocked with 150 items across 14 categories. We tested four conditions (25 trials each): Cart-mounted, Wearable, Dynamic Obstacles, and Sparse Semantics (25-50% of objects randomly removed).

TABLE I
GLOBAL LOCALIZATION IN CONTROLLED ENVIRONMENT

Method	Success (%)	Time (s)	G-RMSE (m/rad)	Tracking (%)
MCL	15%	**1.06**	0.41/0.27	10%
AMCL	10%	17.41	**0.34/0.05**	10%
FQL	90%	4.07	0.42/0.23	59%
DS-MCL	**97%**	1.07	0.36/0.28	**66%**

As shown in Table I, geometric methods collapse due to aliasing. DS-MCL achieves 97% global localization success and a 66% tracking rate, vastly outperforming the deterministic Fixed Quantity Landmark (FQL) baseline. The gap between global success (97%) and continuous tracking (66%) stems from transient dynamic occlusions and sparse

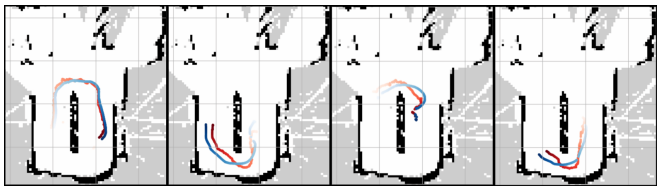

Fig. 3. Qualitative results from the controlled environment showing ground truth (blue) and DS-MCL estimated poses (red). Lighter to darker shading denotes the temporal progression of the trajectories.

semantic signals that occasionally push the pose belief outside the strict tracking threshold. DS-MCL proved resilient to dynamic occlusions and sparse semantics, demonstrating that probabilistic semantic tie-breaking maintains spatial consistency despite active interference. Figure 3 highlights DS-MCL's ability to rapidly converge to the ground-truth trajectory despite ambiguous geometry.

D. Domain B: Large-Scale Operational Facility

To demonstrate zero-shot scalability, we deployed DS-MCL in a 3,500 sq. ft. operational store using Pipeline B. Because instrumenting a public retail space with a motion-capture system is infeasible, ground truth for the evaluation trajectories was established via RTAB-Map SLAM [21].

TABLE II
REAL-WORLD OPERATIONAL FACILITY EVALUATION

Method	Glocal Loc. %	Tracking %	G-RMSE (m/rad)
Real Vision (RGB + Depth)			
MCL	8%	6%	**0.30**/0.25
FQL	42%	38%	0.41/0.09
DS-MCL	**62%**	**40%**	0.34/**0.12**
Upper-Bound (Ground-Truth Semantics, No Depth)			
FQL	82%	70%	0.41/0.31
DS-MCL	**98%**	**92%**	**0.35/0.27**

In this highly aliased environment (Table II), DS-MCL achieved 62% success under real noisy vision. In an upper-bound ablation utilizing flawless ground-truth semantics, DS-MCL achieved 98% success compared to FQL's 82%. This proves that modeling semantics as continuous probability distributions inherently extracts more robust localization signals than fixed-quantity representations.

V. DISCUSSION AND LIMITATIONS

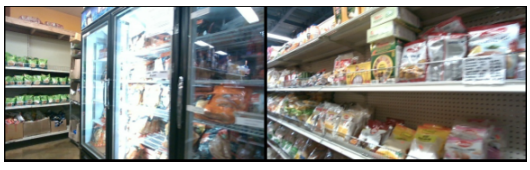

Fig. 4. Perception failure modes in real-world facilities: motion blur, inconsistent lighting, and partial object captures pose challenges for reliable semantic extraction.

While DS-MCL significantly improves localization in quasi-static environments, limitations exist. Like all map-based localizers, we assume an *a priori* map; updating the semantic layer as major inventory overhauls occur is an operational consideration. Second, the system is constrained by perception failure modes (Fig. 4). Sustained perceptual degradation will eventually delay inverse proposals and weaken forward likelihoods.

Furthermore, while framed generally, this robust spatial grounding serves as a critical prerequisite for downstream human-centered applications. In assistive robotics, particularly within navigation aids designed for people with visual impairments, *start-anytime* operation is crucial for shared-control use cases. In these settings, users typically invoke assistance on demand rather than relying on continuous, fully autonomous guidance [22], [23]. DS-MCL's rapid global pose estimation provides the essential infrastructure-free spatial grounding required by such systems to reliably enable other downstream tasks conditioned on robust localization [24].

VI. CONCLUSION

We presented DS-MCL, an approach for vision-based localization in quasi-static environments that models semantics as distributional evidence over object categories. By coupling this representation with an inverse semantic proposal mechanism, the filter remains highly informative under semantic drift and generates targeted global pose hypotheses when geometry is severely aliased. By resolving the aliasing that undermines standard filters, DS-MCL paves the way for more resilient mobile service robots and assistive devices operating in the wild. Future work includes lightweight embedded implementations, integration with downstream navigation stacks, and continuous online maintenance of the semantic layer.

REFERENCES

[1] D. Fox, "KLD-Sampling: Adaptive Particle Filters," in *Advances in Neural Information Processing Systems*, vol. 14. MIT Press, 2001.

[2] C. P. Gharpure and V. A. Kulyukin, "Robot-assisted shopping for the blind: issues in spatial cognition and product selection," *Intelligent Service Robotics*, vol. 1, no. 3, pp. 237–251, July 2008.

[3] N. Zimmerman, L. Wiesmann, T. Guadagnino, T. Läbe, J. Behley, and C. Stachniss, "Robust Onboard Localization in Changing Environments Exploiting Text Spotting," July 2022, arXiv:2203.12647 [cs].

[4] N. Zimmerman, T. Guadagnino, X. Chen, J. Behley, and C. Stachniss, "Long-Term Localization Using Semantic Cues in Floor Plan Maps," *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 176–183, Jan. 2023, conference Name: IEEE Robotics and Automation Letters.

[5] R. G. Goswami, P. V. Amith, J. Hari, A. Dhaygude, P. Krishnamurthy, J. Rizzo, A. Tzes, and F. Khorrami, "Efficient Real-Time Localization in Prior Indoor Maps Using Semantic SLAM," in *2023 9th International Conference on Automation, Robotics and Applications (ICARA)*, Feb. 2023, pp. 299–303, iSSN: 2767-7745.

[6] F. Xie and S. Schwertfeger, "Robust Lifelong Indoor LiDAR Localization Using the Area Graph," *IEEE Robotics and Automation Letters*, Jan. 2024, conference Name: IEEE Robotics and Automation Letters.

[7] "AMCL nav2 documentation," https://docs.nav2.org/configuration/packages/configuring-amcl.html, accessed 2025-09-21.

[8] H. Yin, X. Xu, S. Lu, X. Chen, R. Xiong, S. Shen, C. Stachniss, and Y. Wang, "A Survey on Global LiDAR Localization: Challenges, Advances and Open Problems," *International Journal of Computer Vision*, vol. 132, no. 8, pp. 3139–3171, Aug. 2024.

[9] H. Kuang, X. Chen, T. Guadagnino, N. Zimmerman, J. Behley, and C. Stachniss, "IR-MCL: Implicit Representation-Based Online Global Localization," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, Mar. 2023, conference Name: IEEE Robotics and Automation Letters.

[10] T. Loiseau, G. Bourmaud, and V. Lepetit, "Alligat0r: Pre-training through co-visibility segmentation for relative camera pose regression," *arXiv preprint arXiv:2503.07561*, 2025.

[11] S. Dong, S. Wang, S. Liu, L. Cai, Q. Fan, J. Kannala, and Y. Yang, "Reloc3r: Large-scale training of relative camera pose regression for generalizable, fast, and accurate visual localization," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025.

[12] B. Al-Tawil, T. Hempel, A. Abdelrahman, and A. Al-Hamadi, "A review of visual slam for robotics: Evolution, properties, and future applications," *Frontiers in Robotics and AI*, vol. 11, p. 1347985, 2024.

[13] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *IEEE transactions on Robotics*, vol. 23, no. 1, pp. 34–46, 2007.

[14] E. Goldman, R. Herzig, A. Eisenschtat, J. Goldberger, and T. Hassner, "Precise detection in densely packed scenes," in *Proceedings of the CVF conference on computer vision and pattern recognition*, 2019.

[15] P. Koopman, "Bresenham line-drawing algorithm," *Forth Dimensions*, vol. 8, no. 6, pp. 12–16, 1987.

[16] C. H. Walsh and S. Karaman, "Cddt: Fast approximate 2d ray casting for accelerated localization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3677–3684.

[17] X. Liu, Z. Zhang, J. Peterson, and S. Chandra, "The effect of lidar data density on dem accuracy," in *Proceedings of the 17th International Congress on Modelling and Simulation (MODSIM07)*. Modelling and Simulation Society of Australia and New Zealand, 2007.

[18] S. Y. Kim and K. Cho, "Usability and design guidelines of smart canes for users with visual impairments," *international Journal of Design*, vol. 7, no. 1, 2013.

[19] K. Lee, D. Sato, S. Asakawa, H. Kacorri, and C. Asakawa, "Pedestrian detection with wearable cameras for the blind: A two-way perspective," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[20] A. K. Sengupta and B. Das, "Maximum reach envelope for the seated and standing male and female for industrial workstation design," *Ergonomics*, vol. 43, no. 9, pp. 1390–1404, 2000.

[21] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of field robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[22] Y. Zhang, Z. Li, H. Guo, L. Wang, Q. Chen, W. Jiang, M. Fan, G. Zhou, and J. Gong, "" i am the follower, also the boss": Exploring different levels of autonomy and machine forms of guiding robots for the visually impaired," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–22.

[23] R. Kamikubo, S. Kayukawa, Y. Kaniwa, A. Wang, H. Kacorri, H. Takagi, and C. Asakawa, "Beyond omakase: Designing shared control for navigation robots with blind people," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*.

[24] S. Agrawal, S. Nayak, A. Naik, and B. Hayes, "ShelfHelp: Empowering Humans to Perform Vision-Independent Manipulation Tasks with a Socially Assistive Robotic Cane," in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, May 2023, pp. 1514–1523.