Supplementary Material: Environment Modeling for Service Robots From a Task Execution Perspective

Ying Zhang, *Senior Member, IEEE*, Guohui Tian, *Member, IEEE*, Cuihua Zhang, *Member, IEEE*,
Changchun Hua, *Fellow, IEEE*, Weili Ding, and Choon Ki Ahn, *Senior Member, IEEE*

SUMMARY OF REPRESENTATIVE WORKS ON ROBOT TASK-EXECUTION-ORIENTED ENVIRONMENT MODELING

Category	**Model**	**Reference**	**Core Ideas and Feature**	**Experiment and Evaluation**			
				Sensor Type	**Datasets**	**Real Scenario**	**Real-time Performance**
Localization	2-D model	Yilmaz *et al*. [34]	Using the particle filter localization with the ellipse-based energy model	LRFs (SICK)	/	Static environment	CPU real-time
		Wang *et al*. [35]	Using the LRF-based 2-D occupancy grid map for robot localization	LRF	/	Indoor environment	CPU real-time
		Krajn´ık *et al*. [36]	Based on a spatio-temporal occupancy grid map using a life-long mapping to learn the long-term environment dynamics	LRF	/	Changing environment	CPU real-time
		Kim *et al*. [38]	Feature scan matching of the environment surrounded by glass walls based on the reflection characteristics of laser beams	LRF (SICK)	/	Glass-walled environment	CPU real-time
		Zhang *et al*. [41]	Using 2-D grid map created based on 2-D laser data and pseudo-laser data converted from visual data	LRF (SICK) & vision (Orbbec Astra)	/	Dynamic environment	CPU real-time
		Lee *et al*. [42]	Estimation of robot positions using line features and vanishing point	Monocular vision	RAWSEEDS	Home environment	CPU real-time
		Da Silva *et al*. [43]	By converting RGBD images and combining the descriptive power of CNNs to estimate the robot position	Kinect	Self-built datasets	/	GPU real-time
	3-D model	Fu *et al*. [44]	Based on 3-D point cloud using a combination of point and line information	Kinect	TUM	Office scene	GPU real-time
		Mur-Artal *et al*. [45]	Using a survival of the fittest strategy to generate a compact and trackable map for lifelong operation	RGB-D cameras	NewCollege, TUM, KITTI	/	CPU/GPU real-time
		Rico *et al*. [47]	Using the Monte Carlo technology to realize robot localization based on 2-D grid map and Octomap	2-D LRF, 3-D lasers, RGB-D camera	/	Laboratory, campus	GPU real-time
		Yu *et al*. [48]	Extracting the complementary information from the ceiling and the ground to build a 3-D environment model for the robust localization of the robot	2-D and 3-D LiDARs	/	Home-like environment	GPU real-time
		Marder-Eppstein *et al*. [49]	Using a voxel-based 3-D mapping for robot localization	Two LRFs	/	Office environment	CPU real-time
		Zhen *et al*. [50]	Estimation of robot localization in 3-D maps based on error-state Kalman filter and Gaussian particle filter	Rotating LRF (Hokuyo)	/	Office area	CPU real-time
	Semantic model	Premebida *et al*. [51]	Using dynamic Bayesian network for robot semantic place classification of 2-D maps	LRF	IDOL, COLD-Saarb	/	N/A
		Rosa *et al*. [52]	Leveraging the coexistence of robots and humans to enhance semantic understanding of shared environments	Kinect	/	Domestic environment	N/A
		Balaska *et al*. [54]	Using an unsupervised technique to semantically identify the robot's location in an unknown environment	LRF & Vision	COLD, KTH-IDOL2, New College	/	GPU real-time
		Narayana *et al*. [56]	Lifelong mapping by learning and maintaining semantic information in dynamic environments	Vision	Self-built datasets	/	N/A
		Gomez *et al*. [57]	Constructing the semantic model based on pixel level correlation of depth and semantic data for robot localization	RGB-D sensor (Asus Xtion)	Witham Wharf	Indoor environments	GPU real-time
Navigation	2-D model	Biswas *et al*. [60]	Long-term robot navigation using hierarchical planner based on corrected gradient refinement algorithm	LRF (Hokuyo), Kinect	/	Office environment	CPU real-time
		Zhang *et al*. [62]	User preference-aware navigation via defined virtual area with the common (non-expert) users	LRF (SICK)	/	Indoor environment	CPU real-time
		Song *et al*. [63]	Autonomous navigation in compliant motion based on 2-D map	LRF	/	Laboratory environment	CPU real-time

		Mora *et al*. [67]	Using 3-D data to build a 2-D map representing the geometry of objects based on recursive Bayesian filter	3-D vision	/	Office environment	CPU real-time
		Zug *et al*. [69]	Extracting horizontal scans from Kinect output to simulate laser scanners for robot autonomous navigation	Kinect	/	Office environment	CPU real-time
		Luo *et al*. [71]	Using multi-sensor fusion method with 2-D map for robot indoor navigation	LRF & stereo camera	/	Office building environment.	CPU real-time
		Baltzakis *et al*. [72]	Vector field histogram method modified by fusing laser and vision data for robot navigation tasks	LRF (SICK) & cameras	/	Corridor environment	CPU real-time
	3-D model	Lau *et al*. [73]	Robotic 3-D navigation through incremental updates to changing scenes	Kinect	/	Dynamic environment	GPU real-time
		Gregorio *et al*. [74]	Environment models with multiple representations such as 3-D voxel meshes, 2.5-D height maps and 2-D occupancy meshes for robot navigation	RGB-D cameras (i.e., RealSense)	Freiburg Campus, New College	/	GPU real-time
		Sathyamoorthy *et al*. [75]	3-D environment model composed of multiple stacked 2-D grid maps for autonomous navigation of robots	LRF (Velodyne VLP)	/	Indoor/outdoor scenarios	GPU real-time
		Kim *et al*. [78]	3-D renderable neural radiance map for robot navigation by adopting the particle filter	RGB-D camera	ScanNet	/	GPU real-time
		Maier *et al*. [80]	Using the in-vehicle consumer-grade depth cameras for obstacle mapping, path planning and robot navigation in 3-D environments	RGB-D camera (Xtion Pro Live)	/	Static and dynamic indoor environments	GPU real-time
	Topological model	Blochliger *et al*. [83]	Transformation of sparse feature-based maps from visual SLAM system into 3-D topological maps for robot navigation	Visual-inertial sensor	Self-built datasets	Semi-structured indoor environment	CPU real-time
		Gomez *et al*. [84]	Representation of the environment as a global topology and 3-D dense subgraph using standard CPU	Depth camera (Asus Xtion)	/	Office environment	CPU real-time
		Fraundorfer *et al*. [88]	Real-time scalable topological mapping by adding images and maintaining link graph for robot navigation tasks	Digital camera	/	Corridor and office environment	CPU real-time
		Kim *et al*. [89]	Constructing topology semantic map based on landmark incremental memory acquisition framework	Panoramic	Gibson	Real noisy world	N/A
		Delfin *et al*. [90]	Building visual memory-based topological maps for robot navigation based on key images acquired offline by supervised learning	RGB-D camera (ASUS Xtion)	/	Indoor environment	CPU real-time
Manipulation	3-D model	Gervet *et al.* [92]	Using 3-D feature grids with adaptive resolution to 3-D modeling of the robot's workspace	Azure Kinect	/	Desktop scenario	GPU real-time
		Murooka *et al*. [94]	Update of manipulation region occupancy grid maps from contact information by introducing contact sensor for robot manipulation	Vision and contact sensor	/	Shelf scene	N/A
		Qiu *et al*. [97]	Learning generalizable feature fields based on CLIP feature distillation for robot manipulation	Stereo camera (RealSense D455)	/	Dynamic indoor scene	GPU real-time
		Monica *et al*. [98]	Allowing robots equipped with distance cameras to perform target detection, classification, segmentation and part-based semantic grasping	3D camera (Orbbec Astra-S)	/	Tabletop scenes	GPU real-time
		Terasawa *et al*. [99]	Heuristics-based motion planning for robot manipulation using a combination of sample-based planner and deep learning	Depth camera (Intel RealSense)	/	Bookshelf environment	GPU real-time
	Integrated 3-D model	Wu *et al*. [102]	Integrating 3-D point clouds, topological map and object map to build object-oriented environment modeling	Visual sensor	TUM	manipulation scene	GPU real-time
		Sunderhauf *et al.* [10]	Object-oriented semantic mapping using feature-based RGB-D SLAM, image-based object detection and 3-D unsupervised segmentation	RGB-D camera (PrimeSense)	/	Desk, office, kitchen environments	GPU real-time
		Ruiz-Sarmiento *et al.* [106]	Building a multiversal semantic map for robot manipulation relying on CRFs and contextual relationships	RGB-D cameras	Robot@Home	/	N/A
		Gregorio *et al*. [74]	Environment model with 3-D voxel meshes, 2.5-D height maps and 2-D occupancy meshes for robot manipulation	RGB-D cameras (i.e., RealSense)	Freiburg Campus, New College	/	GPU real-time

		Hornung *et al*. [109]	Combining an octree-based 3-D representation with an anytime search-based motion planner to address 3-D path planning for mobile manipulation	Stereo cameras	/	Cluttered office environment	N/A
	Semantic model	Zhang *et al*. [13]	Safe and efficient mobile manipulation combining grid-based 2-D map, octree-based on local real-time 3-D representation, and ontology technology	RGB-D camera (Orbbec Astra)	/	Indoor environment	CPU real-time
		Gunther *et al*. [111]	Creating semantic environmental model in an incremental and closed-loop manner using a series of 3-D point clouds captured by RGB-D cameras	Kinect	Self-built datasets (Seminar Room, Office)	/	GPU real-time
		Riazuelo *et al*. [113]	Combining ontologies and maps to create knowledge-based semantic models	Kinect	/	Hospital room	GPU real-time
		Bi *et al*. [115]	Creating task-oriented environment semantic model based on knowledge base and environment semantics	RGB-D camera	/	Office environment	GPU real-time
		Rogers *et al*. [116]	Joint reasoning on object and room classification using contextual cues via CRF to provide robots with the capabilities needed to perform various service tasks	Vision (Nikon D90)	/	Indoor environment	N/A
Long-term autonomy	Combined model	Kunze *et al*. [117]	Predicting and inferring robot actions and their effects through combined ontology and simulation techniques	Stereo camera	/	Making pancakes scenario	GPU real-time
		Schenck *et al*. [118]	Using fully convolutional neural network with the integration of temporal information to detect and track liquids to complete the pouring task	RGBD camera	Simulated Dataset Robot Dataset	Pouring task scenario	GPU real-time
		Zhang *et al*. [13]	Integrating 2-D grid map, local real-time 3-D representation and knowledge model to support the robot to perform pick-and-place tasks	RGB-D camera (Orbbec Astra)	/	Indoor environment	GPU real-time
		Hauser *et al*. [120]	Robot performing the task of pushing objects on a flat surface by randomly sampling the robot mode transition	Stereo vision	/	Pushing a block	N/A
		Inceoglu *et al*. [121]	Generating and maintaining accurate environment models of continuous scenes using visual perception and processing methods for robotic tasks	RGB-D camera (Asus Xtion Pro Live)	/	Tabletop scenarios	N/A
	Consistent model	Coradeschi *et al*. [127]	Creating the consistent model that is consistent with the symbolic and physical world to support robotic tasks	Vision	/	Office, surveillance	N/A
		Zhang *et al*. [123]	Building a semantic grounding architecture based on typical spatial location relationships to guide robots to quickly locate task-related objects	RGB-D camera	/	Dynamic home-like environment	CPU real-time
		Persson *et al*. [128]	Semantic consistent modelling using probabilistic reasoning and object anchoring based on rich continuous attribute values measured in sensor data	Kinect2	Self-built datasets	Exemplifying scenarios	GPU real-time
		Elfring *et al*. [129]	Creating a semantic environmental model by introducing probabilistic multi-hypothesis anchoring for robots performing housekeeping tasks	Camera	/	Dynamic scenarios	GPU real-time
		Anderson *et al*. [131]	Allowing robots to perform vision and language tasks by interpreting vision and language as visually grounded sequence-to-sequence translation	RGB-D camera	Self-built datasets	/	GPU real-time
	Probabilistic model	Chikhalikar *et al*. [132]	Using the object probabilistic relationships and contextual priors to search objects based on user preferences	RGB-D camera (Azure Kinect)	/	Indoor environment	GPU real-time
		Aydemir *et al*. [134]	Active visual target search in unknown environments using hierarchical planning based on decision theory and heuristics	Camera	/	Unknown rooms	N/A
		Murali *et al*. [135]	Using a target-driven probabilistic method to reason about the unseen object parts for robot object manipulation	RealSense	/	Cluttered environment	GPU real-time
		Zeng *et al*. [138]	Robot active object search based on a semantic linking map and the spatial probabilistic relationships between objects	RGB-D camera (Primesense Carmine)	/	Home-like environment	N/A
		Honerkamp *et al*. [139]	Combining a large language model and an object-centered environmental model for high-level reasoning to perform household tasks	RGB-D camera	/	Home-like environment	GPU real-time