MASt3R-Nav: WayPixel Navigation in Relative 3D Maps

Vansh Garg[†1], Rohit Jayanti[*1], Krish Pandya[*1], Sarthak Chittawar[*1],
Siddharth Tourani[2,3], Muhammad Haris Khan[3], Sourav Garg[‡1], Madhava Krishna[‡1]

[1] UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386 [2] INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD [3] MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

[†]Corresponding Author, [*]Equal Contribution, [‡]Equal Advising

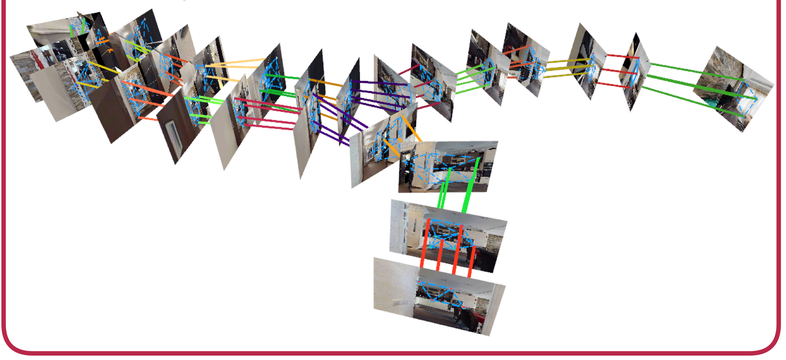
ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2026 VIENNA

mast3r-nav.github.io

Overview: Pixel-Relative Navigation

Problem Statement: Current visual navigation relies on either globally-consistent 3D maps (computationally heavy, requires perfect registration) or object/image topological graphs (lacks geometric precision, leading to poor control).

Our Solution: We introduce a pixel-relative 3D map representation, by leveraging MASt3R's 3D-grounded matching, we build maps based on relative pixel connectivity without needing global consistency.



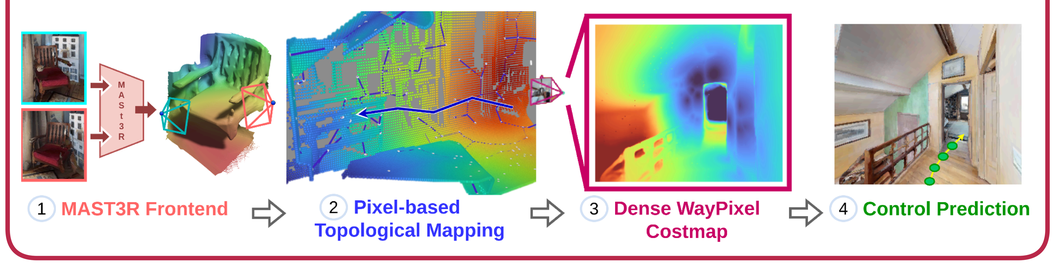
Quantitative Results on HM3D

Method	Type	Train Data	Alt Goal SPL	Alt Goal SSPL	Shortcut SPL	Shortcut SSPL	Reverse SPL	Reverse SSPL	*Average* SPL	*Average* SSPL
GNM	Image-Relative	Real	8.70	15.44	15.38	31.74	3.33	6.11	26.55	34.56
GNM (HM3D)	Image-Relative	HM3D	0.00	10.91	15.38	24.57	13.28	20.77	27.62	35.66
PixNav	Object-Relative	HM3D	26.09	31.66	7.69	22.29	16.16	25.56	23.09	31.57
RoboHop	Object-Relative	Zero Shot	30.43	38.23	30.77	40.87	9.98	16.92	32.19	40.25
ObjectReact	Object-Relative	HM3D	21.74	26.68	23.08	39.64	30.00	42.01	33.36	44.71
MASt3R-Nav (**Ours**)	Pixel-Relative	HM3D	47.83	58.06	46.15	61.10	23.25	26.83	52.79	60.24

Blue cells mark the best scores; Orange cells mark the second-best.

State-of-the-Art Performance: MASt3R-Nav drastically outperforms both image-relative (GNM) and object-relative (ObjectReact) baselines on primary start-to-goal tasks.

MASt3R-Nav: Method at a glance

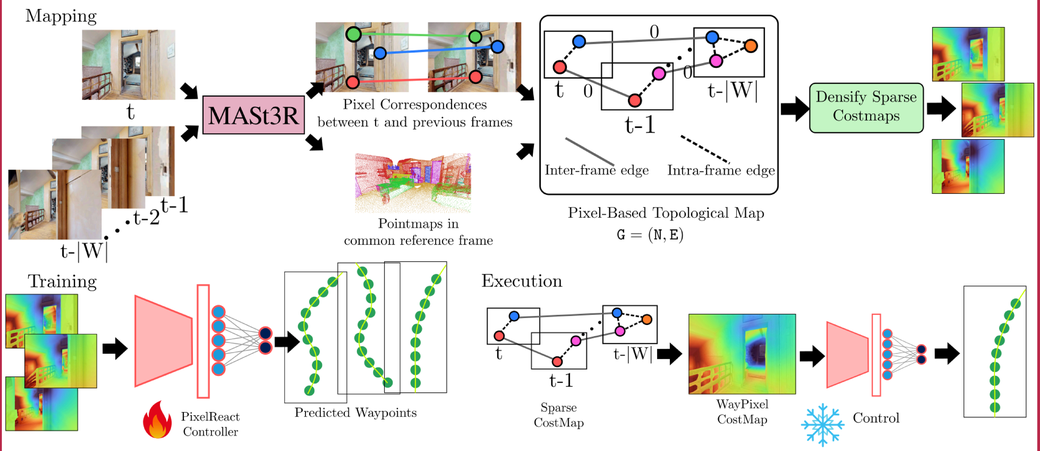


From Pixel Matches to Control Prediction

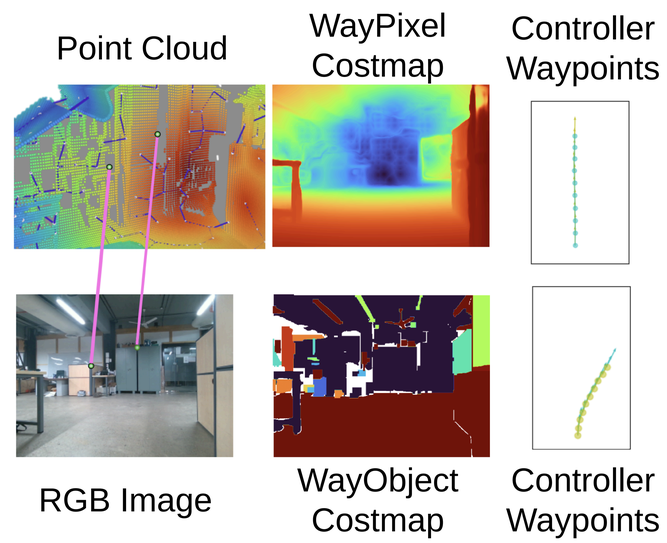
Offline Mapping: MASt3R extracts dense pixel correspondences across sequential images. These form a topological graph where inter-frame edges have zero cost, and intra-frame edges use 3D Euclidean distances.

Online Execution: The query image is matched against the map to localize goal-relevant pixels. These sparse matches are propagated through local 3D geometry to generate a dense WayPixel Costmap.

PixelReact Control: A learned controller is conditioned on the dense costmap to predict a sequence of local waypoints.

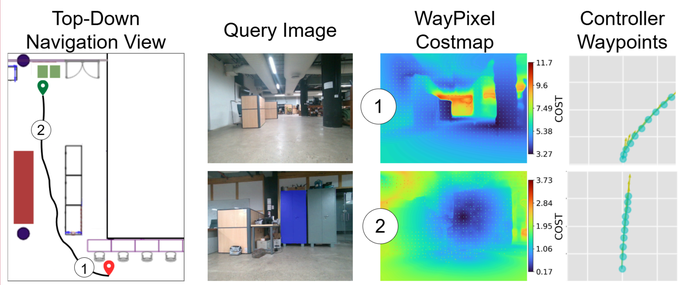


WayPixel v/s WayObject Representation



Our pixel-relative map representation produces dense WayPixel Costmaps that enable the PixelReact controller to correctly navigate to the target ('metal cabinet'). In contrast, coarse object-relative WayObject Costmaps misguide the robot.

Bridging Sim-to-Real Gap



Real-World Transfer:
MASt3R-Nav successfully transfers from sim to real on a P3DX robot equipped with an Intel RealSense camera, navigating to goals using only WayPixel-based guidance.