





Spatial AI: Dense Visual Mapping



Louis Gallagher, John McDonald

¹ COLLABORATIVE DENSE VISUAL SLAM:

- Spatial understanding for multiple collaborating autonomous agents
- Agents work together to build and maintain a shared representation of their workspace



1. Agents independently explore a shared workspace



2. Initially, each agent builds a separate submap of the environment.



² HYBRID SPARSE-DENSE MONOCULAR SLAM:

Dense 3D visual mapping for automotive applications

oop Closur

Combine sparse and dense SLAM approaches with CNN-based dense geometry prediction





2. ORBSLAM-3 performs sparse, feature-based SLAM: suitable for fast, wide-baseline camera motion Feature Extraction

ORB SLAM 3

Tracking:

GBD Camera Trackin







- **3**. A centralised mapping node searches for inter-map loop closures, i.e., overlap between each agents submap
- 4. Overlapping submaps are brought into alignment and fused into a single global map of the environment, shared by all agents

L. Gallagher and J. McDonald. Collaborative Dense SLAM. In Proceedings of the 2018 IPRCS, Irish Machine Vision and Image Processing Conference, 2018, Best Paper Award.

DENSE VISUAL SLAM FOR FISHEYE CAMERAS: 3

Extend hybrid SLAM architecture (2, above) to support wide FOV fisheye cameras within monocular dense SLAM pipeline



1. Fisheye depth estimation CNN builds on PackNet to support Kannala-Brandt camera model without any rectification (Left). Example results on challenging inputs (Right).



4. RGB-D frames are 3. The sparse pose is brought into tight alignment with the dense fused into a dense surfel model via a 6-DOF optimisation based on a joint photometric and 2.5D model geometric cost function



Loop closure constraints generated by ORBSLAM are reflected in the dense surfel map via a deformation graph of affine transforms

L. Gallagher, V. Kumar, S. Yogamani, and J. McDonald. A Hybrid Sparse-Dense Monocular SLAM System for Autonomous Driving. In 2021 European Conference on Mobile Robots (ECMR), pages 1-8, 2021

4 NERF-XR:

• Photorealistic novel view synthesis for near-field, surround-view visualisation in automotive applications



Volume Rendering Rendering Loss – g.t.

(d)



2. Can we use this representation to improve over existing visualization approach based on fixed bowl geometry which results in visualization artefacts?



1. NeRFs represent the radiance field of a scene in the weights of an MLP. Weights are optimized via differentiable rendering

3. We use both KITTI-360 fisheye cameras to construct a surround view model in the periphery of the car



(c)

Reconstruction

using one

camera



Sparse point cloud overlaid on dense metric depth

Per-frame $s_t = \frac{med(D_t)}{med(D_{orb})}$ scale ratio

Sparse map overlaid on color *image for reference*

 $\{s_i \in \mathbf{S} | s_i \in [\mu_{\mathbf{S}} \pm 2\sigma_{\mathbf{S}}]\}$ Rolling buffer to remove noise

2. Scale ambiguous ORBSLAM map needs to be aligned to the metric-scale of the predicted depth maps. We do so by estimating the scale ratio between the predicted dense depth map and a sparse depth map rendered from ORBSLAM, averaged over a rolling buffer of frames to account for scale drift

L. Gallagher, G. Sistu, J. Horgan, and J. B. McDonald. A system for dense monocular mapping with a fisheye camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pages 6478–6486, June 2023





Surround-view NeRFacto Training W/ KITTI-360

