Representations and Computation Patterns for Spatial AI

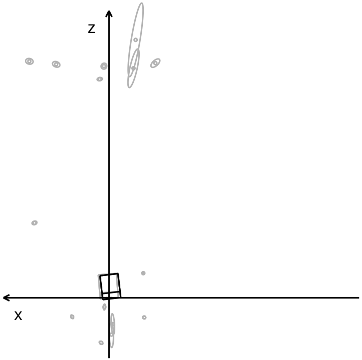
Andrew Davison

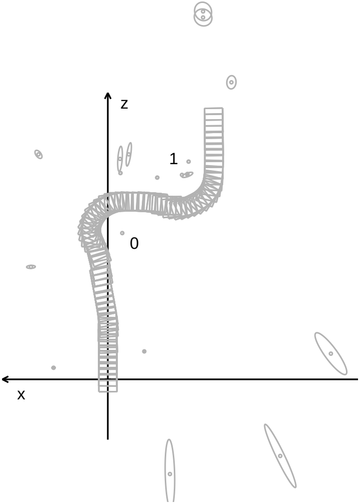
September 9, 2026

SLAM Using Active Vision

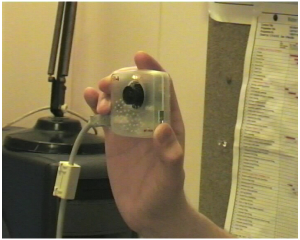
- 1994–1997, Active Vision Lab, Oxford (Davison and Murray, ECCV 1998).
- Persistent sparse visual mapping via probabilistic filtering.

SLAM Using Active Stereo Vision





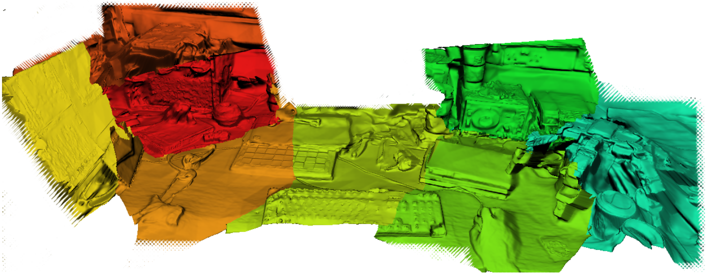
MonoSLAM: Sparse Feature-Based SLAM (Oxford, 2003)



Sparse single camera SLAM focused on egomotion tracking

- Solid 30FPS performance on a 2003 laptop.
- Davison, ICCV 2003; Davison, Molton, Stasse, Reid, PAMI 2007.
- Early demos included AR, wearables and humanoids.
- Improved on by PTAM, Klein and Murray, RRG, ISMAR 2007.

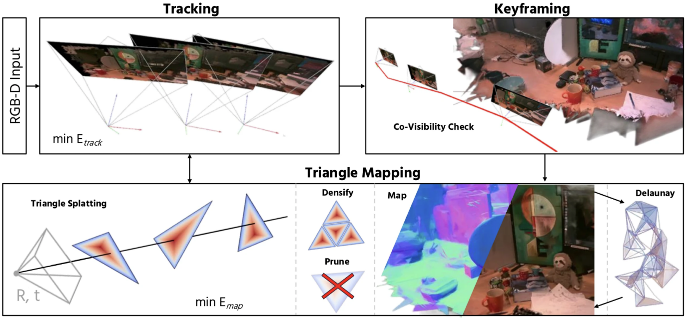
Dense Visual SLAM



Real-time dense reconstruction, and direct camera tracking

- Newcombe *et al.*'s 'Live Dense Reconstruction with a Moving Camera', CVPR 2010 and DTAM, ICCV 2011.
- KinectFusion (Newcombe and Microsoft Research) leads to many other dense SLAM systems.

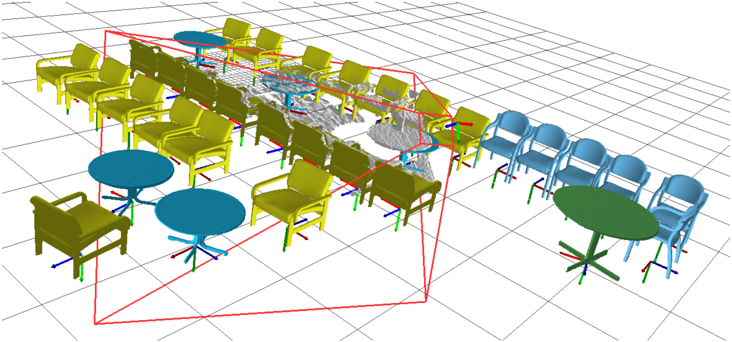
Triangle Splatting SLAM



A map of differentiable triangles

- Fry, Dexheimer, Mazur, Kelly, Davison, ECCV 2026.
- On-the-fly mesh extraction.

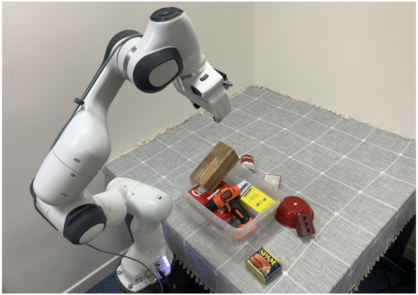
Object-Oriented SLAM



SLAM++: Real-time object level dense SLAM with CAD models

- Salas-Moreno *et al.*, CVPR 2013: bring object recognition to the front of SLAM, and directly build a map at that level.
- Highly accurate and complete but efficient representation.

MoreFusion/ReOrientBot



Multi-object Reasoning for 6D Pose Estimation from Volumetric Fusion

- Wada, Sucar, James, Davison, CVPR 2020.
- CAD models of known objects are jointly fitted to multi-view depth data in difficult situlations with occlusion and contact.
- Learned object reorientation (ReorientBot, ICRA 2022).

SuperPrimitives and 4DPM

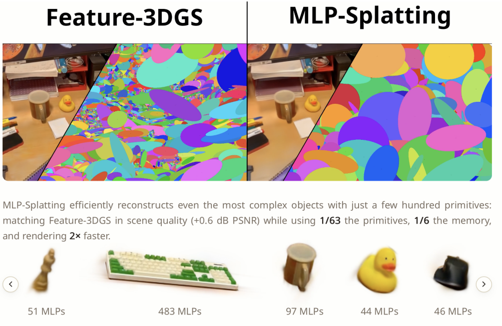
SuperPrimitive, Mazur et al.

4D Primitive Maché, Mazur et al.

Pre-trained networks segment images into small geometry chunks

- Single view learned priors turn a single image into accurate “SuperPrimitives”: object-like chunks.
- Efficient multi-view primitive-level correspondence and optimisation for full 3D reconstruction.
- CVPR 2024, CVPR 2026.

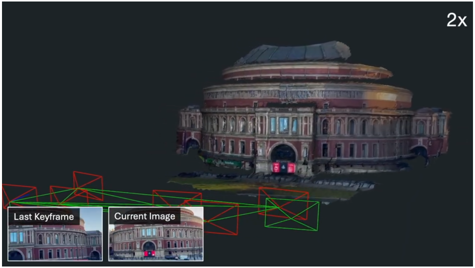
MLP-Splatting



Object-Centric Neural Fields

- Kim*, Cheng*, Kong*, Kelly, Davison, ECCV 2026.
- Primitives composed of local radiance fields centred on anistropic Gaussians.

MASt3R-SLAM

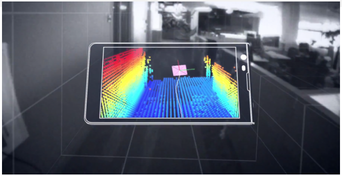


Off-the-shelf MASt3R as strong 2-view prior.

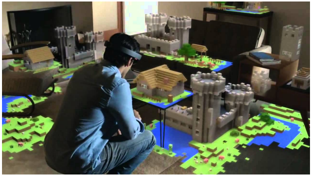
- Similarity pose graph built around 2-view blocks; some optimisation of local point clouds too.
- Works without known camera intrisics (but even better if they are known).
- *Murai, *Dexheimer, Davison, CVPR 2025.

Visual SLAM-Enabled Products

Dyson/iRobot/etc.



ARKit/ARCore/etc.



Meta/Apple/HoloLens/etc.

DJI/Skydio/etc.

Ego-motion and sparse/semi dense reconstruction now mature.

- Many other systems in industrial settings.
- Dense and semantic mapping now rapidly entering early products.

Towards Advanced Embodied Intelligence

Towards World-Changing Applications

- General purpose robotics — a robot to clean and tidy a kitchen?
- Always-on smart glasses (Meta Project Aria Gen2).
- Severe constraints on size, weight, power usage, cost, safety.

Spatial AI: Perception for Embodied Intelligence

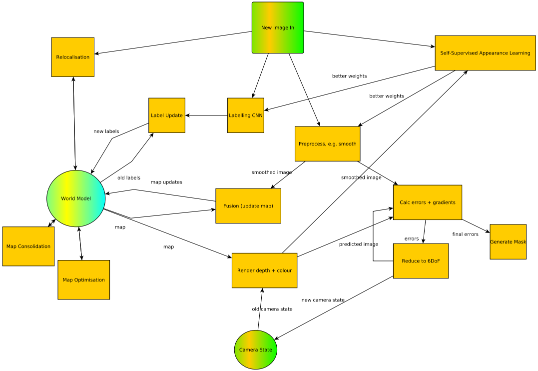
Does embodied AI (robotics) require scene representation at all?

- In principle, an agent could be completely stateless, with a predefined recipe (normally a neural network) for translating current and recent perceptions into action. ‘Pixels to torques!’

What do we mean by scene representation?

- It could be an explicit 3D ‘map’ of the world, with objects, physical and lighting properties, etc. and a model of the robot of interest. Similar to the ‘scene graph’ of a video game.
- Or a purely abstract vector of parameters decoded and updated by neural networks. This idea is recently called a ‘world model’.
- Neither approach has yet enabled general embodied AI.
- Properties we need for real products: extreme efficiency; composability; rapid and complete instantiation; interpretability?

The Computational Structure of Spatial AI



Complicated, loopy storage, computation and communication.

- **This** (and more) is what we need to optimise to get to true Spatial AI. Learned and designed modules.
- Lines of attack: 1. Representation; 2. Hardware.

What will computing hardware be like in the future?

More distributed; local; heterogeneous; especially for embedded efficiency.

- Tiny, cell-like processors with decentralised memory connected in graphs. (Also lower precision, asynchronous.)
- Spatial AI bitter lesson: the algorithms that will win will be the ones that work well on this hardware.
- The most interesting problem for me is the computational structure of a whole Spatial AI system. I think we will see convergence of techniques.

We will presumably laugh one day at today's multi-hundred Watt GPUs, and data centres to train a model that fill a warehouse.

SLAM has always been evolving towards Spatial AI

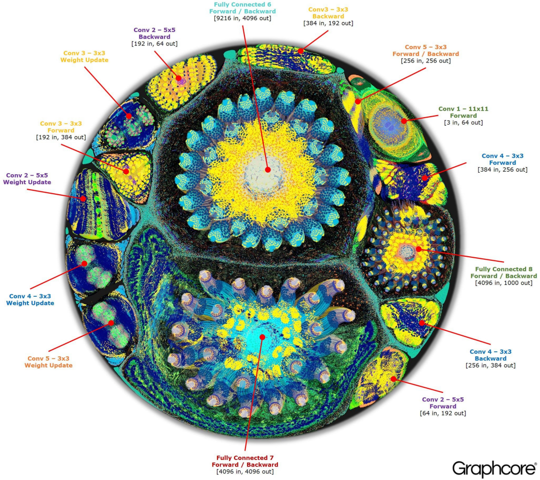
Defining Spatial AI

- Spatial AI is the online problem where vision is to be used, usually alongside other sensors, as part of the Artificial Intelligence (AI) which permits an embodied device to interact usefully with its environment.

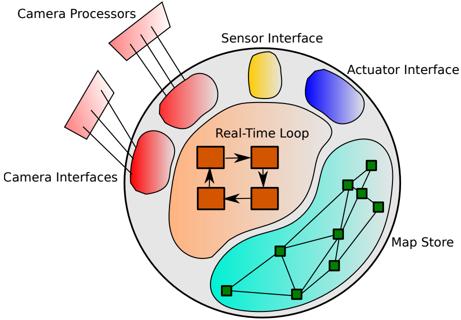
References for more details:

- FutureMapping: The Computational Structure of Spatial AI Systems and FutureMapping 2: Gaussian Belief Propagation for Spatial AI (Davison, Ortiz, arXiv 2018, 2019).
- SLAM Handbook, especially Chapter 18. Cambridge University Press, coming soon, but available free online now.

Visualising a Processing Graph (Graphcore)



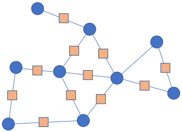
Finding the Graphs in Spatial AI



Can we lay out the main elements of Spatial AI processing 'geographically' on an imagined graph-like processor?

- World model/scene representation internal processing.
- Close-to-sensor and close-to-actuator processing.
- Main 'real-time loop' that connects these together.

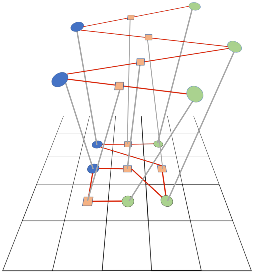
Gaussian Belief Propagation for Spatial AI



Factorised computation via Gaussian Belief Propagation

- **In-place computation** on a loopy graph.
- Each node of a factor graph can be placed on a different computing node; communication by message passing.
- A Visual Introduction to Gaussian Belief Propagation `gaussianbp.github.io`. Ortiz, Evans, Davison, 2021.
- **Distributed convergence**: when GBP converges, it achieves correct marginal means, and marginal covariances which are overconfident in a predictable pattern (e.g. Weiss and Freeman, Neurips 1999).

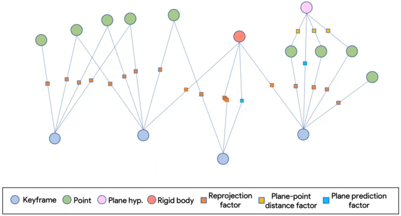
Bundle Adjustment on a Graph Processor



Ortiz, Pupilli, Leutenegger, Davison, CVPR 2020.

- BA implemented with graph node-wise parallelism and solved by GBP message passing.
- Non-linear, robust reprojection factors.
- Implemented on Graphcore's IPU: for some real problems, 30 times speed improvement over Ceres on CPU.

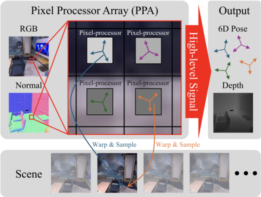
Incremental Abstraction in Distributed Probabilistic SLAM Graphs



Ortiz *et al.*, ICRA 2022.

- Priors/cues which are external or internal to the graph can be used to trigger abstraction hypothesis, which can later become permanent.
- Spatial AI graphs will exist at multiple, hierarchical abstraction levels simultaneously.
- Implemented on Graphcore's IPU.

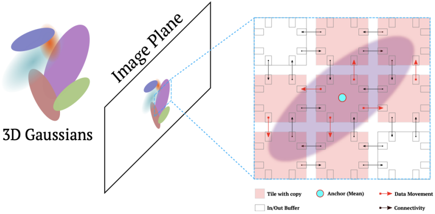
How Much Computer Vision Can we Push to an On-Sensor Processor?

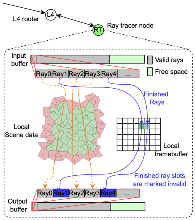


Dense visual odometry and more?

- On Sensor Vision Project: Manchester, Bristol, Imperial.
- In-pixel visual odometry via a hierarchical graph topology: PixVOD: Kim, Alzugaray, Rhodes, Kelly, Davison, ECCV 2026.
- Future: general segmentation and scene flow. What type of connectivity would be ideal?

Rendering 3D Scenes on a Graph Processor





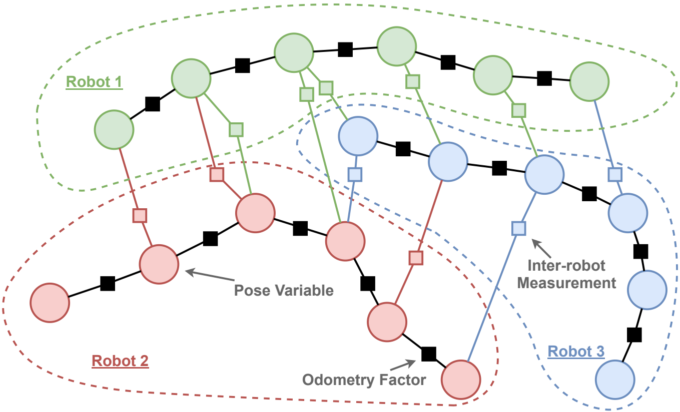
Two approaches to parallelism

1. In the framebuffer (image plane): move 3D Gaussians between tiles associated with image blocks as needed, taking advantage of gradual camera motion. “Rendering 3D Gaussians on a Graph Processor”, Fry, Alzugaray, Pupilli, Kelly, Davison, EGSR 2026.
2. In the 3D scene: lay the 3D elements across tiles; trace rays by passing messages from tile to tile. “Radiant Foam Rendering on a Graph Processor”, Tuya, Alzugaray, Fry, Davison, arXiv 2026.

Interoperable, Heterogeneous Multi Robot Systems?

- Many different companies will make robots and devices. How to make them interoperate, safely and efficiently?
- Is a unified cloud 'maps' system needed? Or is there a viable decentralised alternative?

Robot Web: Core Design



- Divide up the full factor graph between robots.
- GBP on full factor graph. Inter-robot message passing via asynchronous web interface.

Spatial AI Conclusions

Spatial AI Research

- Persistent scene representations, built from a device's own sensors, are crucial for the most general types of embodied AI.
- What is the right representation that enables efficient, composable mental simulation?
- The computational patterns of representations and algorithms, and how they relate to processing and sensing hardware, are the most important research issues for the coming years.

Affiliations (`https://x.com/ajddavison`)

- Imperial College London, Department of Computing.
- Slamcore, applied Spatial AI startup in London.