Title: Mono-Hydra: Real-time 3D scene graph from monocular input with an IMU using a UAV system

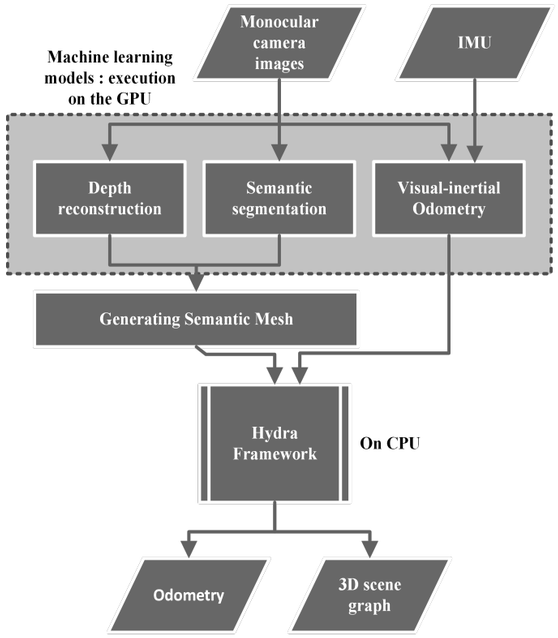
PRESENTER:
Bavantha Udugama
b.udugama@utwente.nl

INTRO:

Spatial Perception system

- Implemented on Real-time Drone system
- Representing spatial relationships in 5 layers of a hierarchical graph structure
- Simple monocular camera input with Deep learning-based Depth and Semantic prediction.

METHODS



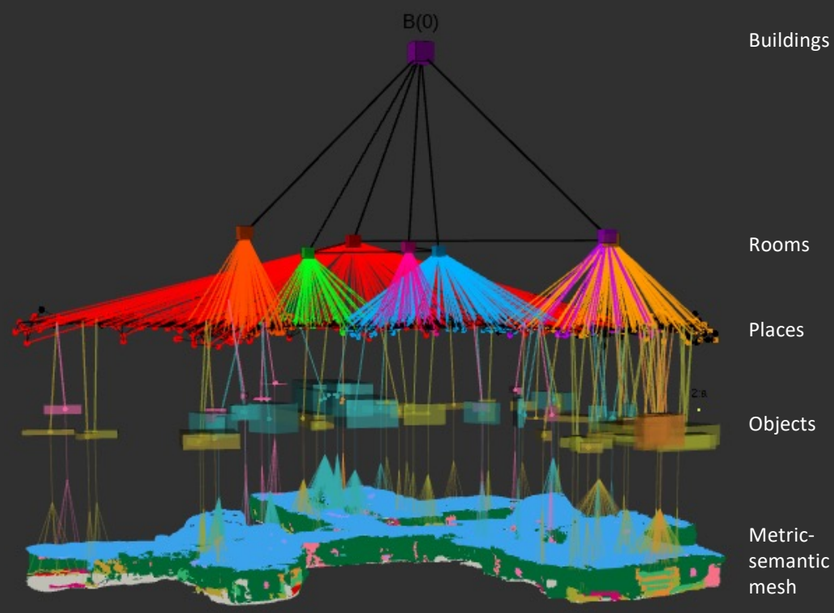
RESULTS

- Real-time semantic mesh generation with a monocular camera
  a. Depth prediction with RMSE 4.561 (better than the state-of-the-art 4.594 in Monoformer)
  b. Semantic Segmentation with HRNet with mIoU 80.4 in indoor scenarios.
- Deriving 3D scene graphs in real-time only with CPU (10 Hz)

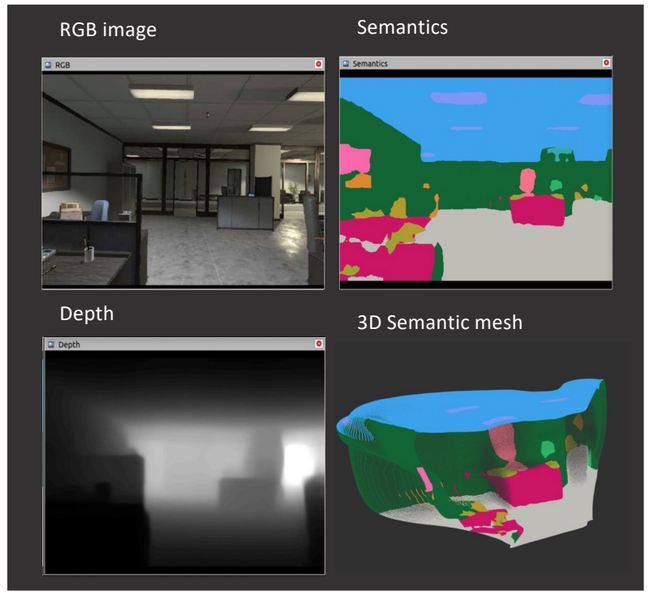
CONCLUSION:

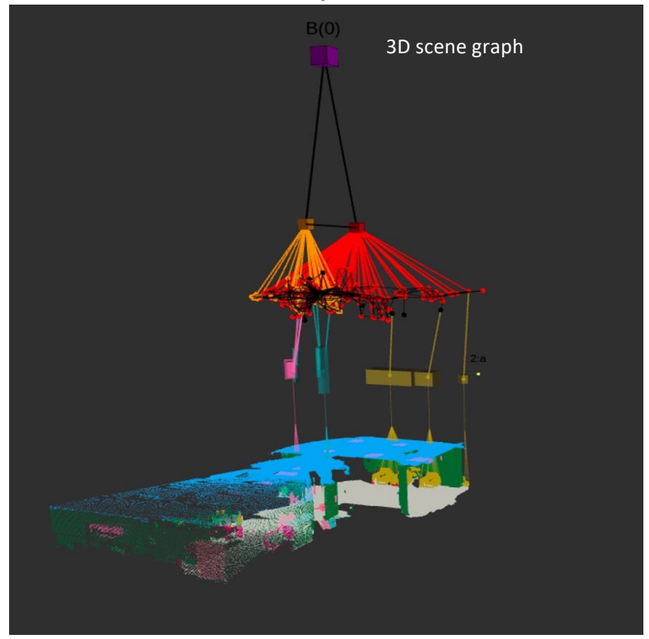
The proposed Deep learning framework with the Hydra framework can utilise a simple camera to generate 3D scene graphs in real time.

Advanced Mental Model for robots; Human-like **spatial understanding** is possible with a single **camera in real-time**.



Scan QR for more details





DISCUSSION

- How to embed dynamics of the environment into a 3d scene graph
- Will combined networks for depth and semantic prediction improve semantic 3D mesh generation?
- Possibilities of utilising geometric features in 3D scene graphs to optimise deep learning networks.
- Can we label rooms like “office room”, “kitchen”, “living room”

Prof. Francesco Nex, Prof. George Vosselman

UNIVERSITY OF TWENTE. | ITC