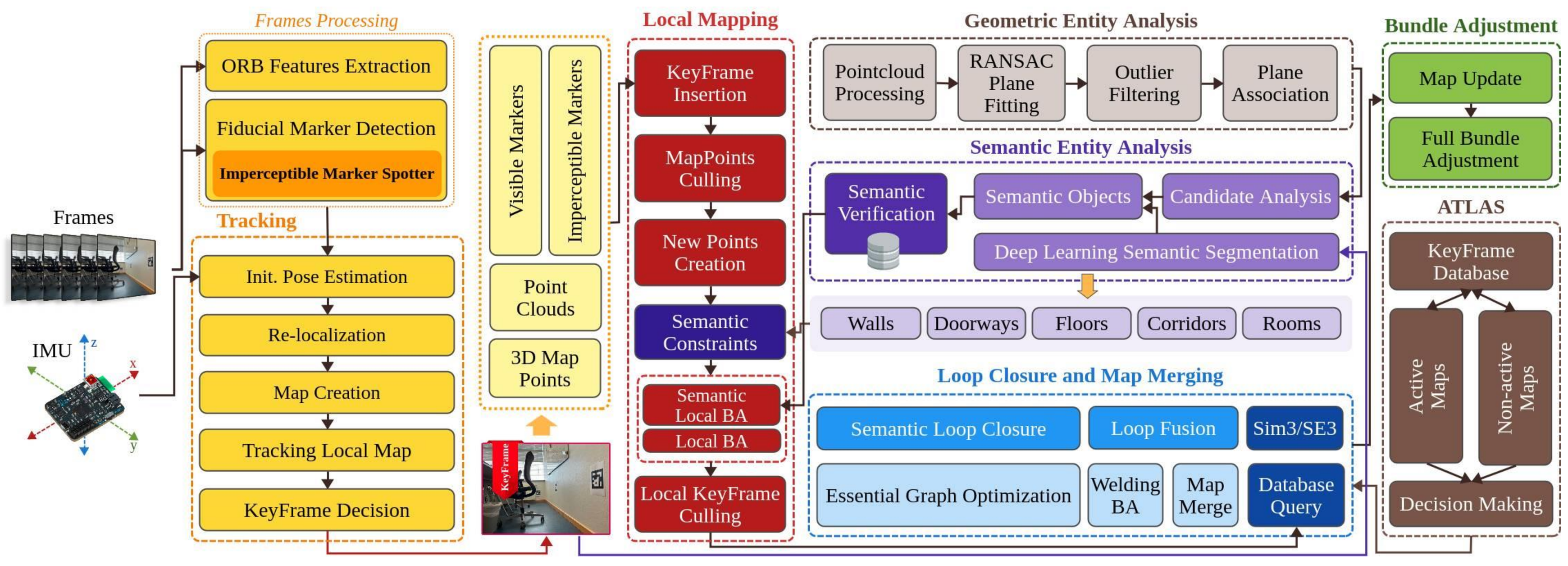
Late Breaking Results on vS-Graphs: Integrating Visual SLAM and Situational Graphs for Multi-level Scene Understanding

uni.lu | SnT

Ali Tourani[1], Hriday Bavle[1], Saad Ejaz[1], David Morilla-Cabello[2], Jose Luis Sanchez-Lopez[1], and Holger Voos[1]

[1] Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, Luxembourg
[2] Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain



Introduction and Motivation

Providing high-level **Scene Understanding** using situational graphs [6] and building them from **visual data** enables reasoning about:

- **Geometry**, as in traditional SLAM solutions (*e.g.,* ORB-SLAM 3.0 [4]),
- **Semantics** from a Deep Neural Network (DNN) semantic segmentation module,
- **Knowledge augmentation** through labeling various structural-level objects using ArUco markers [1],
- And cheap and widely available **sensors**: robots, phones, VR/AR devices, etc.

Problem Statement and Solution

Incorporate different sources of **visual knowledge** (geometric, semantic, and marker-based) for mapping higher-level semantic entities, including **rooms** and **corridors** [3]. What we combine in a situational graph are:

- Geometric mapping (ORB-SLAM3 / Keypoints / Dense PointCloud)
- **Semantic mapping** using a DNN (pFCN)
- **Marker-based** information (the idea taken from our previous work in [3], which was based on UcoSLAM [2]).

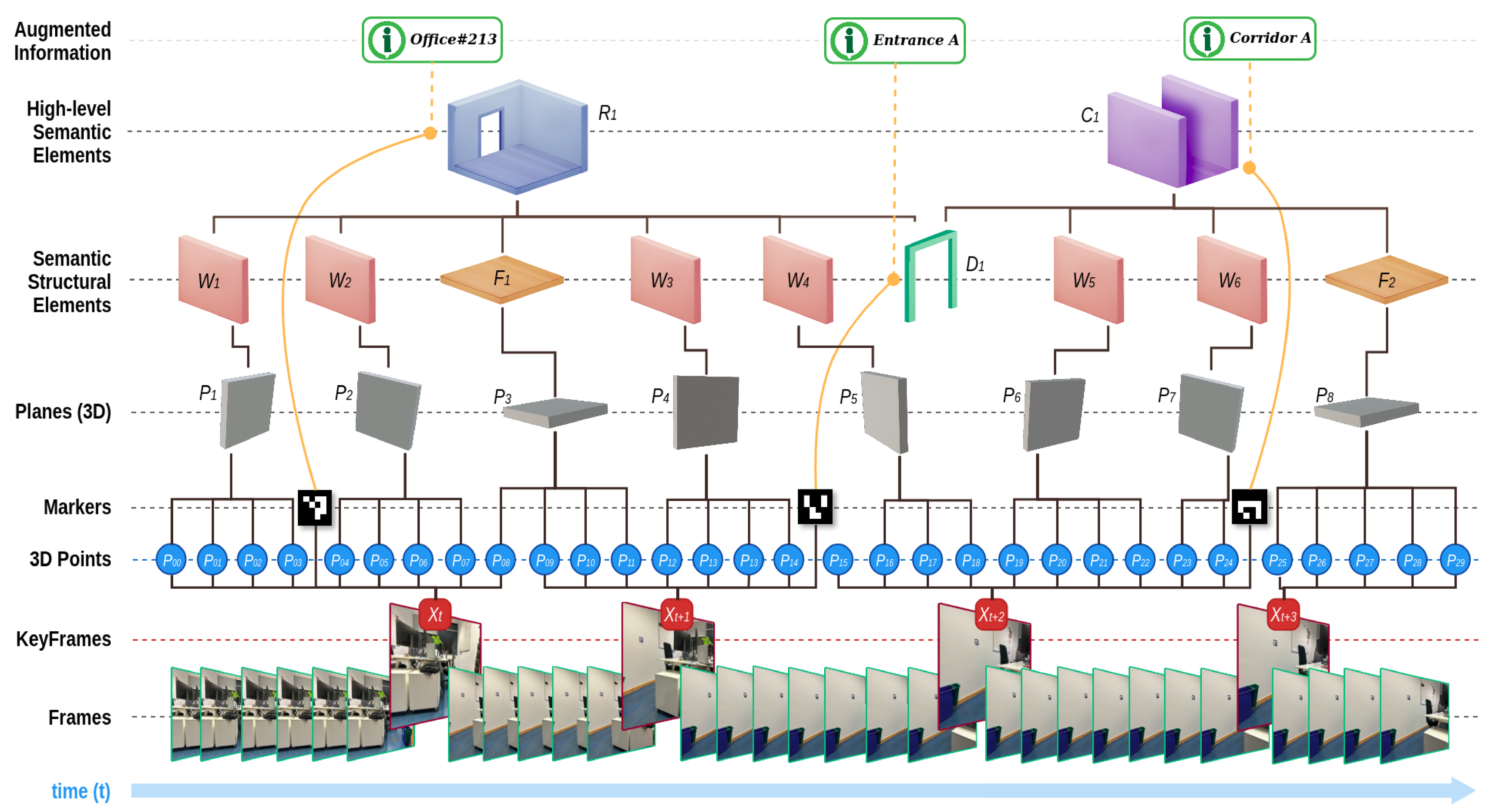
For richer map building, we incorporate:

- More robust and precise **structural extraction** with geometric and semantic information, as an improvement over [5],
- Integration of **knowledge augmentation** through ARUCO markers,
- Integration in an **end-to-end framework**,
- **Real-time** performance of all the modules.

ICRA2024 YOKOHAMA | JAPAN

The base idea [5]

Our IROS'23 Paper [3]



Architecture Design

Baseline: ORB-SLAM 3.0.

New threads added to the baseline:

- Geometric Entity Analysis: plane based RANSAC
- Semantic Entity Analysis: semantic object detection and matching, respectively.

Processes:

1. **PointClouds** obtained from RGB-D / Mono-SLAM are passed to Geometric Analysis to fetch all the 3D planes and their equations without knowing their type,
2. In Semantic Analysis, the portion of the **point clouds** related to walls and grounds are matched with the previously detected planes to add semantic information to them. Detection of walls and corridors are done based on the geometrical layouts of structural-level entities.

References

[1] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion," Pattern Recognition, vol. 47, no. 6, pp. 2280–2292, 2014.

[2] R. Muñoz-Salinas and R. Medina-Carnicer, "Ucoslam: Simultaneous Localization and Mapping by Fusion of Keypoints and Squared Planar Markers," Pattern Recognition, vol. 101, p. 107193, 2020.

[3] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, R. Muñoz-Salinas, and H. Voos, "Marker-based Visual SLAM Leveraging Hierarchical Representations," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, US, pp. 3461-3467, 2023.

[4] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An Accurate Open-source Library for Visual, Visual-Inertial, and Multimap SLAM," IEEE Transactions on Robotics, vol. 37, no. 6, pp. 1874–1890, 2021.

[5] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, D. Isinsu Avsar, R. Muñoz-Salinas, and H. Voos, "Vision-based Situational Graphs Generating Optimizable 3D Scene Representations," arXiv preprint arXiv:2309.10461, 2023.

[6] H. Bavle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, "S-graphs+: Real-time Localization and Mapping Leveraging Hierarchical Representations," IEEE Robotics and Automation Letters, vol. 8, no. 8, pp. 4927-4934, 2022.

We are hiring! More details: **joseluis.sanchezlopez@uni.lu**