Dense 3D Reconstruction and Extrinsic Calibration

Aditya Dhawale, Kumar Shaurya Shankar, and Nathan Michael

Objectives:

- Perform extrinsic calibration of sensors of different modalities (RGB, depth, motion capture)
- Render a dense and accurate representation of an environment using Bundle Adjustment (BA)

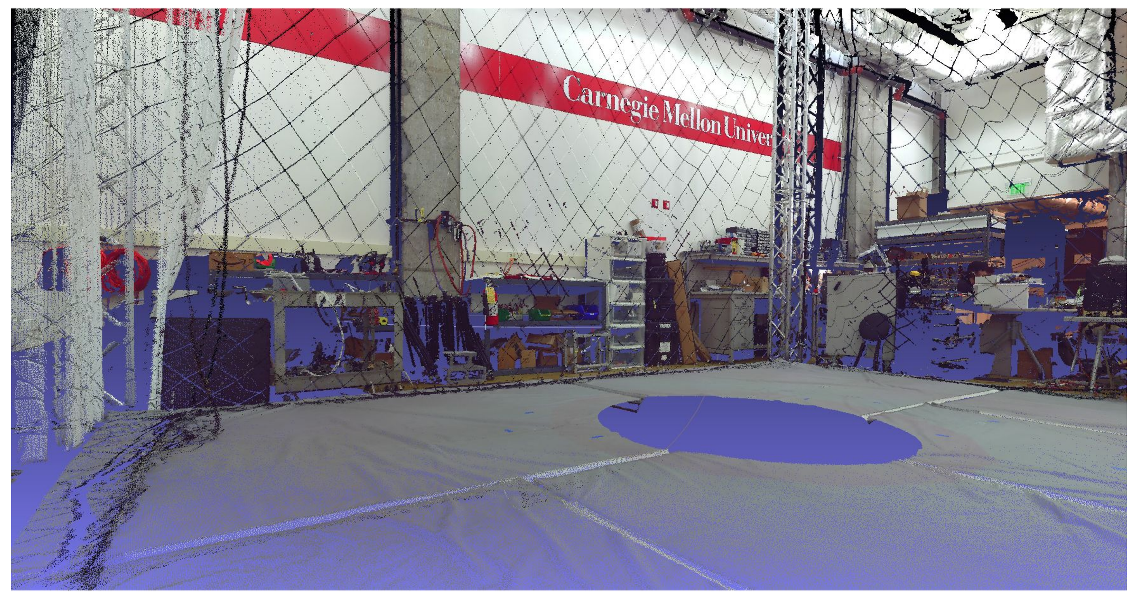

Fig. 1. 3D reconstruction using a precision laser survey system.

Challenges:

- Requirement of a framework capable of indoor and outdoor operations
- State-of-the-art sensor systems e.g. precision laser survey systems (FARO scanner) are expensive and require stationary operation (Fig. 1).

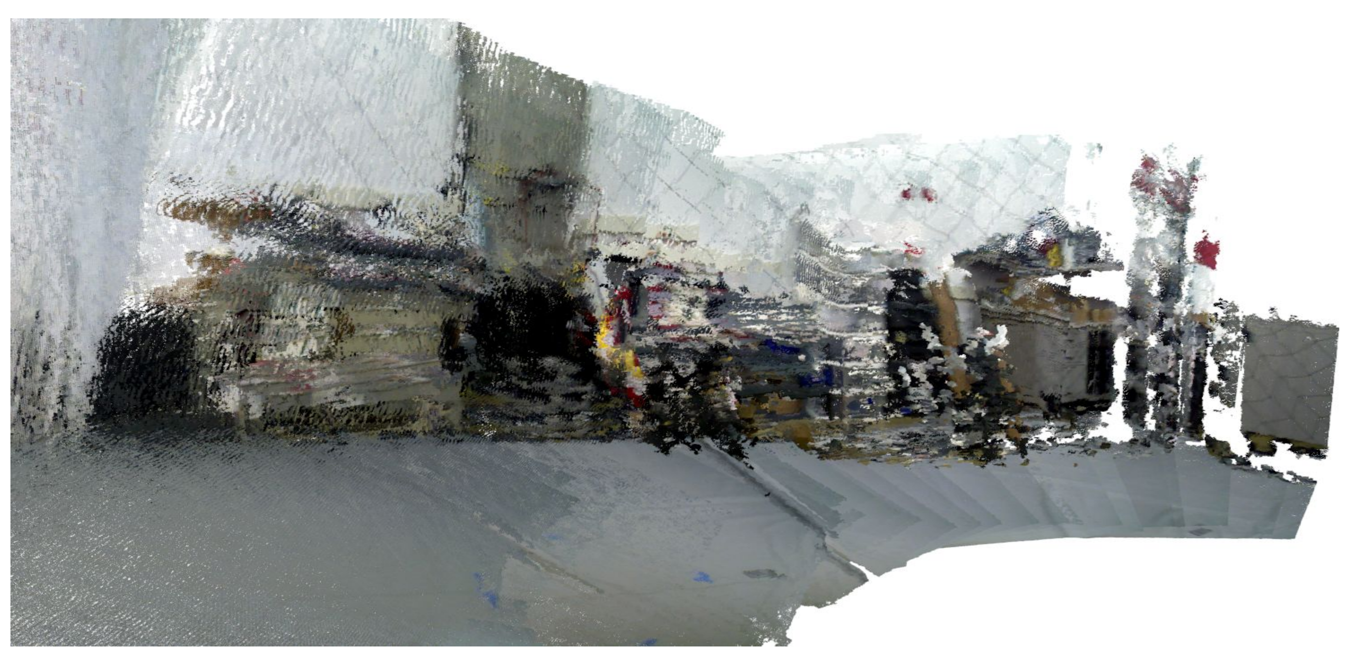

Fig. 2. Noisy 3D reconstruction due to the drift in odometry data (visual monocular odometry or motion capture odometry).

References:

- R Unnikrishnan, M Hebert, *"Fast Extrinsic Calibration of a Laser Range Finder to a Camera"*, Technical Report, 2005, *Robotics Institute, Carnegie Mellon University.*
- Kyel Ok, W. Green, and N. Roy, *"Simultaneous Tracking and Rendering: Real time Monocular Localization for MAVs"*, in *Proc. of IEEE International Conference of Robotics and Automation*, May 2016, pp. 4522-4530.

Multi-Modal Sensor Calibration:

- Extrinsic calibration between a depth sensor (e.g. IR camera, LIDAR) with another sensor (e.g. RGB camera, motion capture) performed by planar segmentation of the calibration plane from depth scan and comparing it with the same calibration plane in the other sensor's frame using Singular Value Decomposition (Fig. 3-4).
- Intrinsic calibration of the cameras performed with OpenCV fisheye calibration routine.

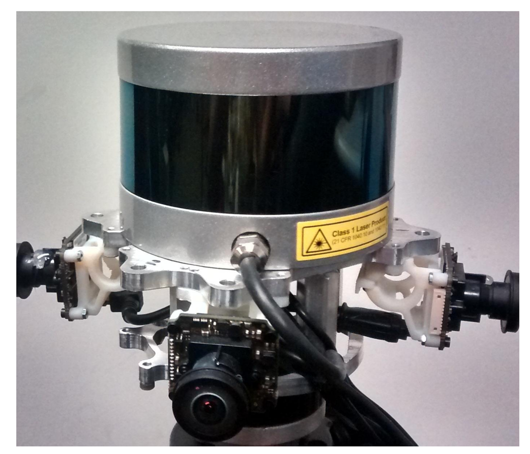

Fig. 3. Sensor rig with four RGB cameras and a Velodyne LIDAR.

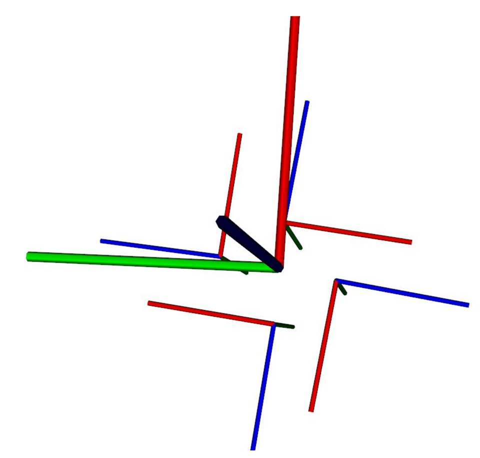

Fig. 4. 3D representation of the extrinsic calibration of sensor rig (Fig.3).

Method	Number of Keyframes Selected	Final Reprojection Error (pixels)	Total Time (sec)
Hybrid	110	8.95 * 10^11	1158
Euclidean Distance	79	3.54 * 10^12	830
Coverage Ratio	70	3.43 * 10^12	2543
Time Interval	47	3.47 * 10^12	153

Table 1. Comparison of L2 norm of reprojection error for different keyframe selection heuristics.

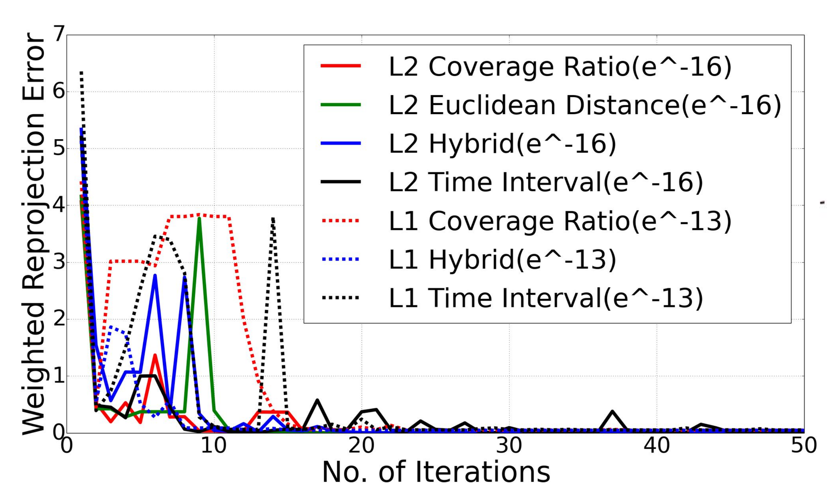

Fig. 5. Comparison of L1 and L2 norm of reprojection error for different keyframe selection heuristics.

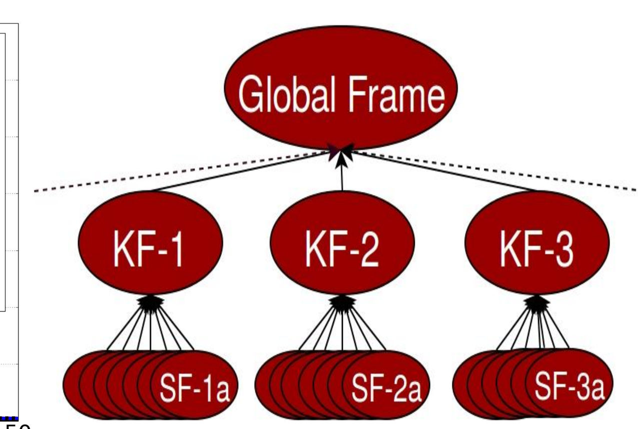

Fig. 6. Hierarchical structure of keyframe based BA.

Keyframe Based Bundle Adjustment:

- Minimization of weighted reprojection error
- Odometry data used as information prior
- Implementation of Iterative Closest Point (ICP) algorithm based on 2D SIFT feature matching
- Keyframe selection heuristics :
  - Fixed time difference : encodes temporal changes
  - Fixed Euclidean distance : encodes rotational and translational changes
  - Keyframe image area coverage ratio : encodes the number of matched features
  - Hybrid : Combination of all the above heuristics

Reprojection Error for keyframe block k,

$$R_k = \sum_{i=0}^{I} \sum_{n=0}^{N_{i,k}} w_n \left\| \mathbf{x}_{i,n} - \pi(T_i^k \pi^{-1}(\mathbf{x}_{k,n}; \delta)) \right\|_2$$

where,

I is the total number of sub-frames in the k^{th} block,

$N_{i,k}$ is the total number of features matched in the i^{th} sub-frame image of k^{th} block and k^{th} keyframe image,

w_n is suitable weighting factor which reduces with increasing depth,

$x_{i,n}$ is the pixel location of a matched feature n in i^{th} sub-frame image and

$x_{k,n}$ is the corresponding feature location in the k^{th} keyframe image,

π is the intrinsic camera matrix,

T_i^k is the transformation matrix between the i^{th} sub-frame and k^{th} keyframe,

δ is the depth value at a particular pixel coordinate.

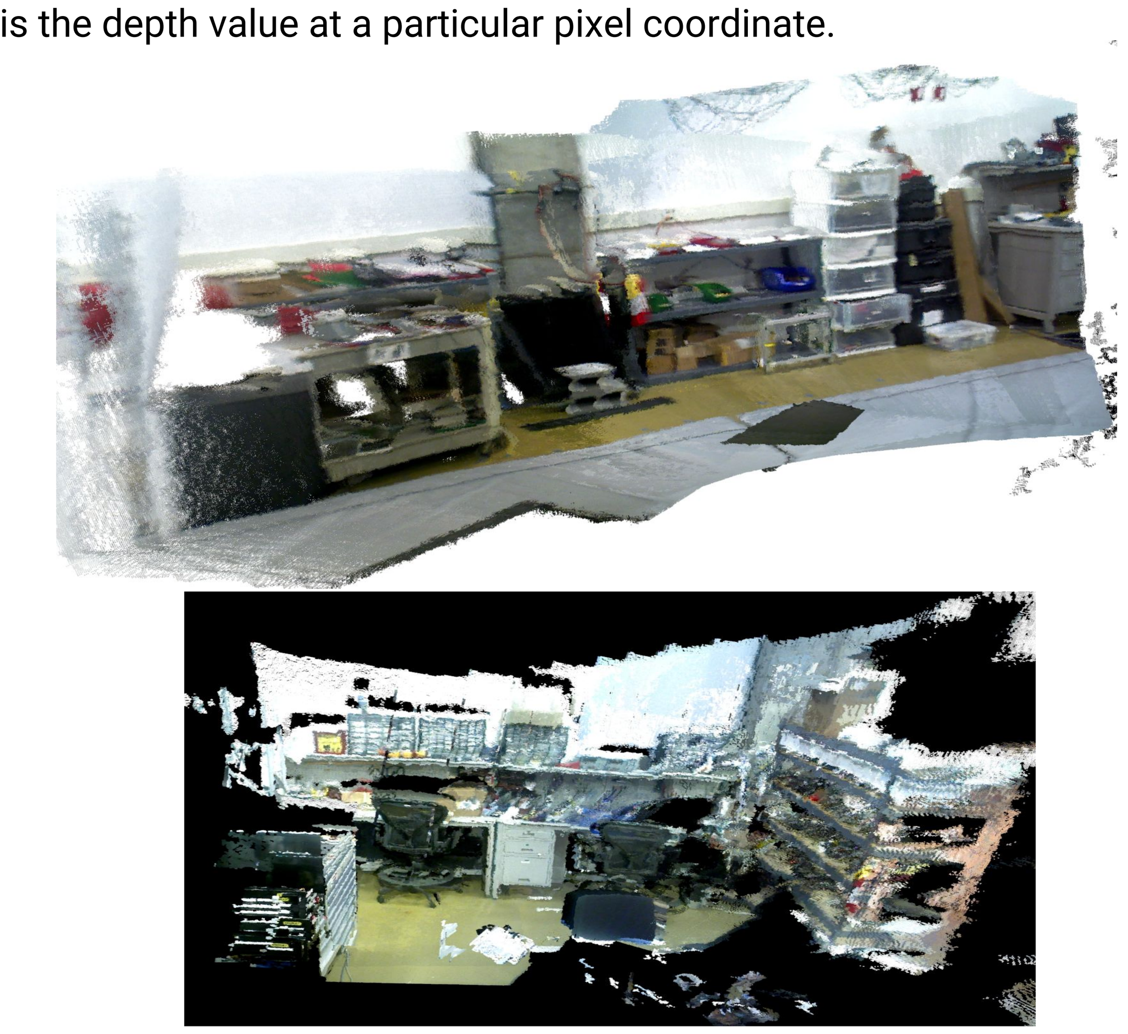

Fig. 7. 3D reconstruction using RGB-D sensor with vicon motion capture odometry data (top) and visual odometry (bottom) as prior.