Multi-floor Indoor Localization for a Human-Operated Backpack

George Chen, Timothy Liu, Matthew Carlberg, John Kua, Avideh Zakhor

Video and Image Processing Lab, University of California, Berkeley

{gchen,timothyliu,carlberg,jkua,avz}@eecs.berkeley.edu

Automated 3D modelling of building interiors is useful in applications such as virtual reality and entertainment. Using a human-operated backpack system equipped with 2D laser scanners and inertial measurement units, we use scan-matching-based algorithms to localize the backpack in complex indoor environments such as a T-shaped corridor intersection, and two indoor hallways from two separate floors connected by a staircase. The localization results are used to (a) generate textured 3D scene models, and (b) enable image based rendering of indoor environments.

We mount orthogonally positioned 2D laser scanners and two inertial measurement units (IMU's) on a backpack, as shown in Figure 1. Orthogonal placement of the laser scanners allows us to run scan matching to recover five backpack pose parameters over time. One IMU is a navigation grade Honeywell HG9900, which provides highly accurate measurements of all six pose parameters and serves as our ground truth. The other IMU is an InterSense InertiaCube3, which provides orientation parameters. We use the laser scanners and the InterSense IMU to localize the backpack.

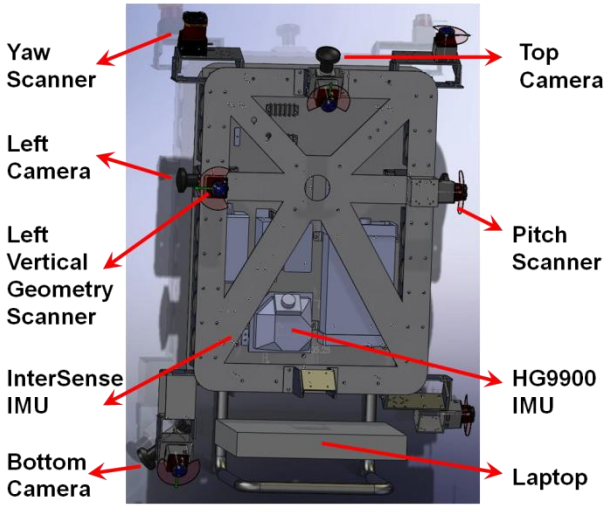

Figure 1: CAD model of backpack system

When the backpack is worn by a human operator, the direction of forward motion is x, leftward motion is y, and upward motion is z. Roll, pitch, and, yaw are defined as rotations around the x, y, and z axes respectively. We use the yaw scanner to estimate x, y, and yaw, and the pitch scanner to estimate z of the backpack pose via scan matching [1]. Lastly, we use the InterSense IMU to estimate roll and pitch. We enforce loop closure by applying the Tree-based Network Optimizer by Grisetti et al [2] to globally optimize our estimated poses, accounting for locations revisited and making use of scan matching and sensor uncertainty.

We test our localization algorithm on two datasets: a T-shaped corridor intersection (set 1), and two indoor hallways from two separate floors connected by a staircase (set 2). Estimated trajectories and associated error characteristics are shown in Figures 2 and 3 respectively. Figure 4 shows a snapshot of the textured 3D model resulting from set 3. In generating this model, we used the vertical scanner on the left side of the backpack to capture geometry, and three cameras to generate texture for the resulting geometry.

We use the localization results to enable virtual walkthroughs using an image based renderer. The renderer uses a three-step process to determine which image to display. First, it locates an initial set of neighbouring camera positions relative to that of the viewer. A dot product between the viewer and camera's orientation vectors provides a threshold to eliminate image planes facing the wrong direction. The renderer chooses the final image from the nearest neighbouring camera. Then the RANSAC algorithm is used on SIFT features from neighbouring images to find an optimal homography to stitch images for an increased field of view. In addition, the localization algorithms can generate plane fitted models for occlusion

detection within the renderer. If an intersection with a plane occurs between two camera positions, the images are occluded and no longer considered to be neighbours. This filters both the initial set of neighbouring images and the set for stitching images together. The image-based renderer performs at 25 frames per second (fps) when one image is rendered and at 5 fps when 4 images are stitched per frame on an unstructured set of 800 images.

[1] G. Chen, J. Kua, S. Shum, N. Naikal, M. Carlberg, and A. Zakhor. "Indoor Localization Algorithms for a Human-Operated Backpack System," to be presented at 3D Data Processing, Visualization, and Transmission 2010, Paris, France, May 2010.

[2] G. Grisetti, C. Stachniss, and W. Burgard. "Non-linear Constraint Network Optimization for Efficient Map Learning," IEEE Transactions on Intelligent Transportation Systems, Volume 10, Issue 3, Pages 428-439, 2009.

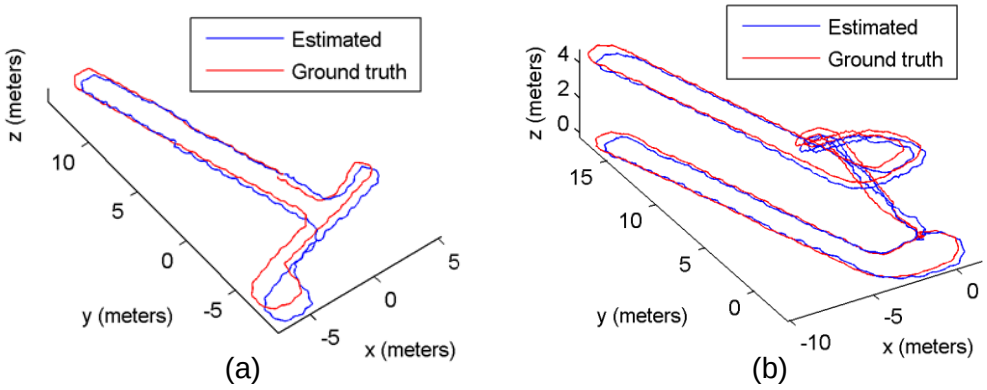

Figure 2: Estimated trajectory vs. ground truth for: (a) set1 and (b) set 2

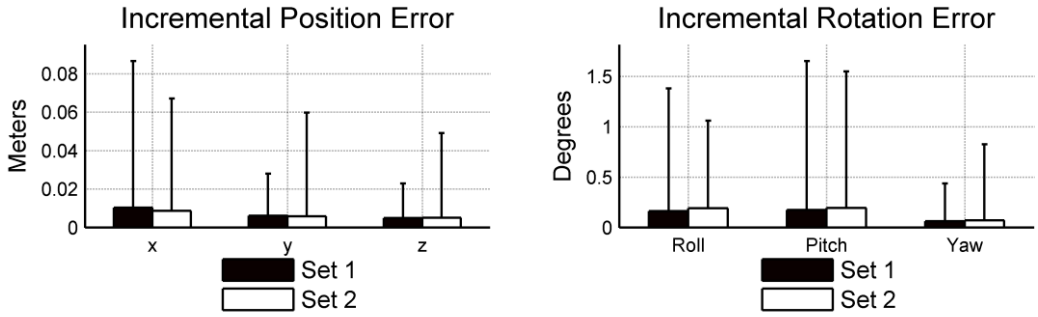

Figure 3: RMS error for estimated poses (lines above bars denote peak errors)

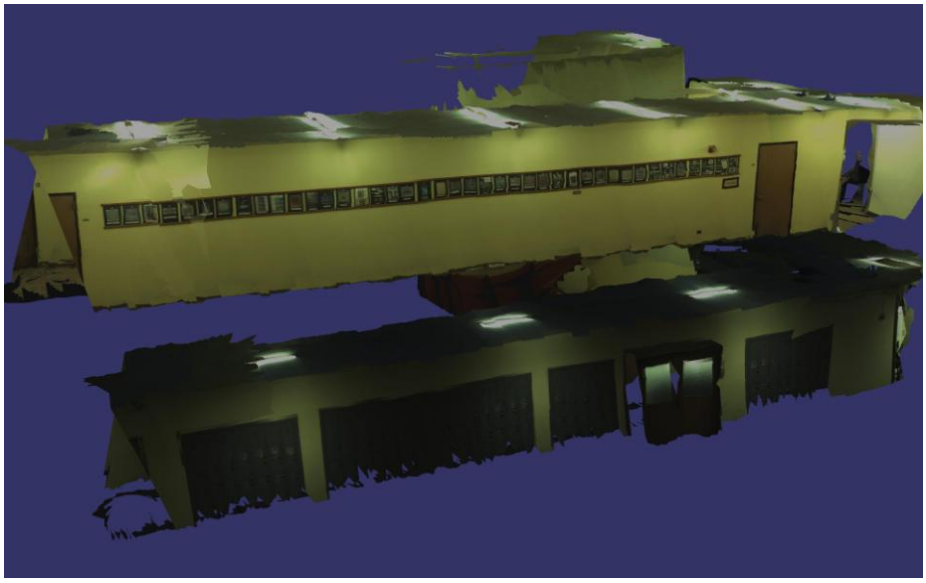

Figure 4: Snapshot of the textured model for set 2