Accurate Fiducial Mapping For Pose Estimation Using Manifold Optimization

Xiao Hu[1], Jakobsen Jakob[1], Knudsen Per[1], Wei Jiang[2]

[1]National Space Institute, Technical University of Denmark
[2]School of Electronic and Information Engineering Beijing Jiaotong University

DTU Space
National Space Institute

Problem

The accurate pose estimation for moving objects within a given workspace is one of the most fundamental tasks for many applications.

A typical solution is to use a motion capture system such as Vicon[a], Optitrack[b] which can provide pose with high precision. However, these systems are usually costly and can only work indoor with limited coverage.

Another common solution is the integration system with high-end Inertial Navigation System (INS) and a Real Time Kinematic (RTK) Global Positioning System (GPS). Nonetheless, its performance will deteriorate dramatically in clustering environments and even stop to work in the GPS denied places.

[a]http://www.vicon.com
[b]http://www.optitrack.com

Contributions

We propose an approach to build a map of fiducial markers based on manifold optimization and then extend the fiducial map for pose estimation. Process pose fusion and optimization directly on manifold to avoid numeric problems, singularities and maintain smoothness. The fiducial map based pose estimation system is cost-effective, lightweight and can work both indoor and outdoor.

Preliminaries

G: the global coordinate system. **M**: the local coordinate system of fiducial marker. The local coordinate system of i^{th} marker is represented by $\mathbf{M}_i$. **C**: the camera coordinate system. Similarly, $\mathbf{C}_k$ denotes the camera coordinate system of the k^{th} image.
$\mathbb{SO}(3)$ is defined with (1):

$$\mathbb{SO}(3) = R \in \mathbb{R}^{3\times3} | RR^T = I_{3\times3},\ det(R) = 1 \tag{1}$$

The Special Euclidean Group $\mathbb{SE}(3)$ is further defined (2):

$$\mathbb{SE}(3) = \left\{ T = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \in \mathbb{R}^{4\times4} \| R \in \mathbb{SO}(3), \mathbf{t} \in \mathbb{R}^3 \right\} \tag{2}$$

An additive increment in the vector space will associate to an multiplication increment on manifold, which follows the below approximation

$$exp((\xi + \delta\xi)^\wedge) \approx exp((\mathcal{J}\delta\xi)^\wedge) exp(\xi^\wedge) \tag{3}$$

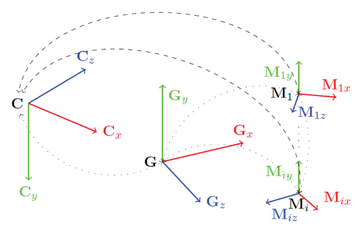

Figure 1: Coordinate Systems. Dashed connections are transformation matrices directly measured and dotted connections are transformation matrices to be optimized.

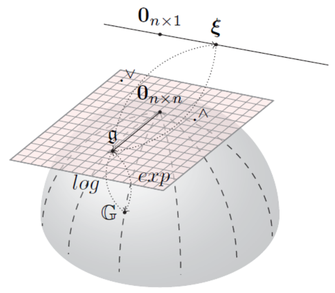

Figure 2: Mapping relationship among matrix Lie Group, Lie Algebra and the vector space.

The pinhole camera model which is appropriate to most cameras is used in this work. We assume that the camera is supposed to be well-calibrated and the intrinsic parameters are precise since no optimization will be made on them.

References

Methods

This section details each step of the proposed method with a systematic chart shown in Fig. 3. ArUco detector implemented in OpenCV [**?**] for fiducial detection and estimate pose with the RPnP algorithm [**?**]. The relative pose from i^{th} marker to j^{th} marker, if both viewed in k^{th} camera frame, can be further computed as

$$T_{\mathbf{M_i}}^{\mathbf{M}_j -1} = T_{\mathbf{M}_j}^{\mathbf{M}_i} = T_{\mathbf{M}_i}^{\mathbf{C}_k -1} T_{\mathbf{M}_j}^{\mathbf{C}_k} \tag{4}$$

By scanning all images collected, an initial pose graph *pg* is incrementally established. A schematic diagram depicting the data structure of *pg* is shown in Fig. 4.

The aim of pose fusion is to refine the estimation of $T_{\mathbf{M}_i}^{\mathbf{M}_j}$ by fusing all validated measurements obtained. After pose fusion process, a refined pose graph which better represents relative transformations of markers will be created. The Gauss-Newton method is used for pose fusion. In order to establish optimal paths, the fused pose graph pg_f is treated as an undirected graph with nodes, edges denoting markers and costs, respectively. Graph search method is used to find the origin which possess the minimum accumulated costs to all other vertices. The pose optimization is to further refine poses by minimizing per frame reprojection errors using the Levenberg–Marquardt (LM) algorithm [**?**].

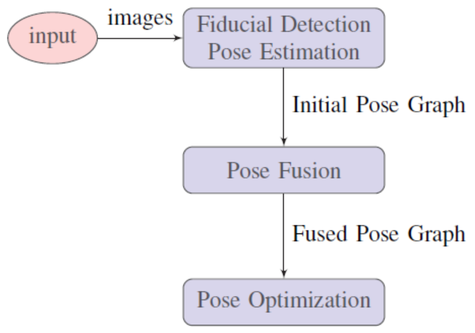

Figure 3: Flowchart of proposed approach.

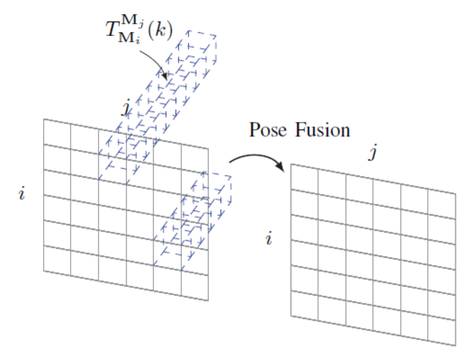

Figure 4: Schematic diagram of the established initial pose graph *pg* and pose graph pg_f after pose fusion process.

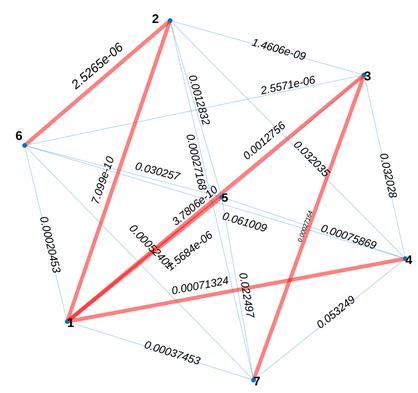

Figure 5: Undirected Graph with blue dots denoting markers, solid light blue edges denoting costs. Red thick edges describe the minimum spanning tree obtained.

Results

Synthetic Mapping Test:

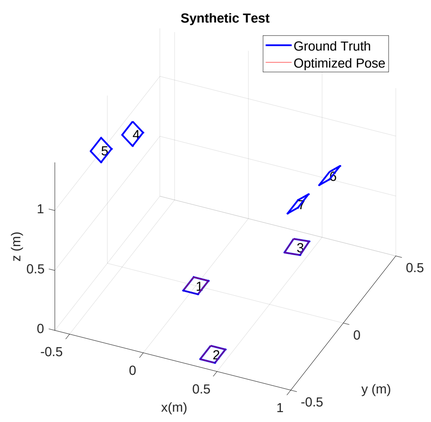

Figure 6: Synthetic Mapping result. Blue squares denote ground truth and red squares denote the optimized markers' pose.

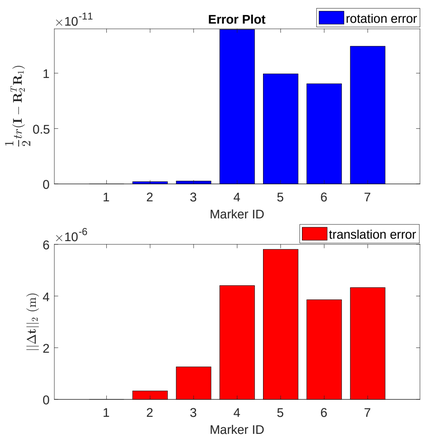

Figure 7: Rotation and translation error analysis. Marker 1 is set as the origin. Consequently, its rotation and translation errors are zero. Translation errors are measured in meters.

Box Test:

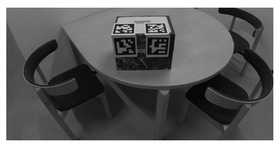

Figure 8: Experiment setup: left is the snapshot of "box" experiment.

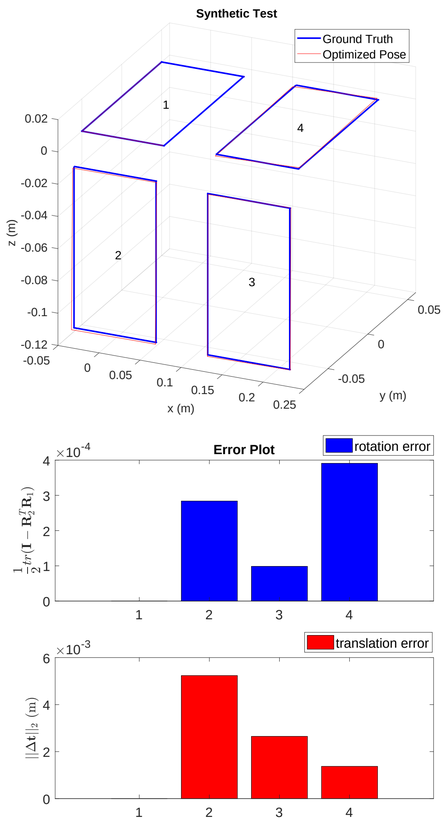

Figure 9: Mapping result and error plots from "box" experiment. Marker 1 is the origin so its erros are zero.

Pose Estimation Test: the GoPro Hero 6 is used to capture images with a resolution of 2.7K (2704 × 1520).

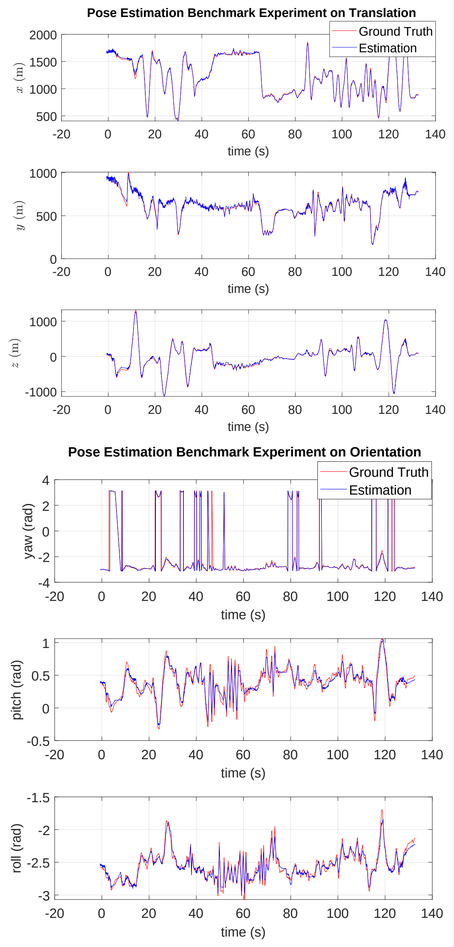

Figure 10: Benchmark Test of pose estimation from fiducial map (blue) and ground truth (red).

Acknowledgement

This work is supported by the Innovation Fund Denmark via project UAV-QMS.