CIS II VR Guided Surgery - Registration Pipelines Background Reading

Ruixing Liang
Johns Hopkins University
rliang7@jh.edu

Hongchao Shu
Johns Hopkins University
hshu4@jhu.edu

Instructors: Max Zhaoshuo Li; Prof. Mathias Unberath; Prof. Russell Taylor

1 Introduction

Real time and simultaneous tracking of both surgeon region of interest(ROI) and surgical tool is of great significance to computer assisted surgery. However, existing tracking solution is disruptive in current surgical workflow especially within the scope of our focus mastoid surgery. An urge to tailor a stereo video-based tracking for accurate surgical navigation system has been developing recently since the burgeon of large scale deep neural network (DNNs). Though researchers are craving for large quantity of good data with expert level of annotations for either semi- or supervised learning, they could never be satisfied simply using existing traditional algorithm based registration pipelines. It involves cascaded calculations and transformations which intuitively amplify the error throughout each step which has been evaluated in the first half of the project process. Therefore, we proposed a novel combination of real world data and generation of virtual reality annotations to refine and expedite the data generation process both in favor of better surgical training and preoperative plan but more importantly for overcoming the data scarcity issue existed wide-spread in DNNs training.

2 Literature Review

Such system have been explored recently by researchers and commercial companies. However most of which are not open-sourced for research and adaptations are not possible. This restricted themselves from developments and expansions of application scenarios. A focus on Adan et al. group's great work in Johns Hopkins since will be given followingly since this was the fundamentals where our works building upon.
They proposed a virtual surgical simulation framework AMBF+ where we could utilize open sourced AMBF Description Format (ADF) to define objects freely with the help of Blender.(See Figure1) Basically, they have demonstrate this framework capabilities by developing a haptic interacted virtual environment specifically for mastoidectomy procedure.

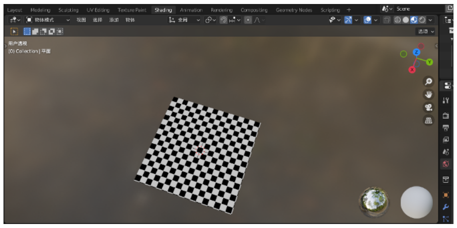

Figure 1: ADF Build in Blender View

They further contribute to one of the most important feature our works will heavily depend on that is the optimized point cloud generation. Adnan improved perspective view non-linear depth map embedded in OpenGL by GPU linearization however this maybe intentionally developed for VR trainning purpose instead of for training DNNs. Because linear depth or smoothed depth may be too ideal which may in turn enlarge the gap between simulation and real scene. This problem has already been identified in authors' conclusion as visual realism as the future work directions. This depth map could be further exploited in generating the mask based on shader. In our case, however, we will use the microscope video stream instead only take the mask, therefore, the drawback could be perfectly mitigated by this mean. Most of their efforts have been made to enabling user complete workflow and experience using AMBF+, for example multiple ways to interact including haptic and VR equipment API support and some extent of flexibility in development.

As a result, our work could potentially become a great completion of AMBF+ framework serving as a additional plugin to incorporate real world stereo cameras and optical tracker as additional peripherals support for library and offer more modalities for down stream applications especially in mastoidectomy procedure. Aside from the general review of our foundation project from Adnan. We also extend our research on background in three main fields including camera calibration, hand eye calibration and sensitivity analysis in following sections.

2.1 Camera Calibration

To date, plenty of camera calibration methods have been proposed, two main streams including self-calibration and conventional ones. Self calibration requires accurate knowledge of motion of cameras, whereas conventional methods used fixed camera, a looser condition for calibration. The crucial component in conventional camera calibration approach is the patterns selections, 3D and 2D patterns are available for calibrations. 3D patterns only requires one shot, a direct linear transformation algorithm is all we need to calibrate the camera. The apparent limitation is that it is expensive to fabricate an accurate 3D micro-scale cube, this will significantly downgrade the calibration result.

We dive into one paper in optics field: *Telecentric stereo micro-vision system: Calibration method and experiments*

So as Chen et al. Decided, planar pattern based calibration should be the most cost effective way to calibrate cameras. A homographic transformation matrix maps 2D mark points of the pattern to 2D image points at each position. Based on the invariance of the camera's intrinsics, a factorization method can be devised to decompose the constraints provided by several homographic matrices. Since we do not have the prior knowledge of the camera lens parameters, in particular whether if it is telecentric lens, we would like to form a more general solution to stereo camera calibrations. Reprojection errors, planar estimation, axis error covariance are certain evaluation metrics, researchers considered to decide calibration is good or bad.

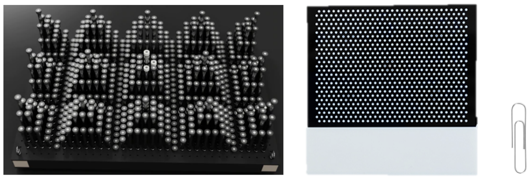

Figure 2: Multiple Commercially Available Patterns

2.2 Hand-Eye Calibration

While high closed-loop accuracy of the surgical navigation system requires almost perfect performance of each part within the system, we decided to perform hand-eye calibration to determine the pose and orientation of camera w.r.t the tracker coordinate instead of keeping on using motion net to inference that information.

As mentioned, our specific goal is to make the closed-loop error within 1mm, which requires the hand-eye calibration method has the highest performance ever. The traditional way to solve the problem is to use the method in [1], first solve for rotation and then translation. While translation part largely relies on the rotation part, error accumulate to translation also. The dual-quaternion method

can simultaneously solve for rotation and translation, which solves the error propagation. And the selected paper used the dual-quaternion method as core solver.
The technical approach of this paper is clearly structured. It mainly contains four part, target extraction, time alignment, hand-eye calibration, and the refinement step.

1. Target extraction
   The author used the targets with known geometry named AprilTag. Detect the corners on the target to find the pixel coordinates, then use RANSAC based Perspective-n-Point method and the camera intrinsic parameters to find the corresponding 3D coordinates of corners w.r.t camera. The author set a threshold λ_{th} to filter out frames with too many outliers.

2. Time alignment
   The sensors of the system might not communicate with each other, thus it's essential to align the timestamp of data from each sensor. The author correlates the angular velocity norms of both pose signals to do this.

3. Hand-eye calibration
   The hand-eye calibration is basically set up as an AX=XB problem using dual-quaternion. Solving this problem with SVD needs to filter out outliers and noise. So the following filtering method are implemented:

   - Pose filtering: This filter first calculate screw motion axis for each dual-quaternion, then exclude pose pairs whose screw axis are almost parallel by setting a threshold of dot products of each combination. This can improve the efficiency of calibration.
   - RANSAC classic: First randomly sample pose pairs and reject those whose scalar part of dual-quaternion are not equal. Then use RANSAC to iteratively choose the inliers agree with the hand-eye calibration, set position and orientation threshold to identify outliers. Repeat the hand-eye calibration on inliers to refine.

4. Refinement step
   Using the global approach above as the initial guess, the author performed a joint maximum likelihood optimization of calibration and trajectory. The author used the Lie group valued B-splines to represent the SO(3) trajectories.

This paper presented a precise, applicable, and well-packaged hand-eye calibration solution, which can be used in many scenarios. It refines hand-eye calibration procedure not only on the hand-eye calibration algorithm, but also focus on other parts within the close-loop, including target extraction, time alignment, data pre-filtering, and nonlinear optimization refinement. The author demonstrates importance of each part with good visualization combined with numerical tables. And it also supplied 3 types of data from different calibration frameworks to evaluate its generalization.
There're still some shortcuts. The toolbox may be too well-packaged, it's relatively hard to use parts of it independently. For our project as an example, we don't need the time alignment implementation in this paper, for we use ROS environment and already got every pose aligned.
Though, the author used three types of frameworks to prove the toolbox has good generalization, it turns out a limited performance on our setup. In target extraction part, it's not always the best choice to use the AprilTag, for our framework, we have to use very small targets (10x20mm) to calibrate the microscope, thus we have to write our own extraction method from scratch.

2.3 Sensitivity Analysis

Uncertainty quantification through computer simulation is an intrinsically interdisciplinary field that has seen a rapid growth in the last decades. Its goal is to find sources of uncertainty in each component of the physical quantities simulation and propagate them into model responses. Many methodologies are included in such a broad definition, including structural reliability, sensitivity analysis, reliability-based design optimization, and Bayesian techniques for computer model calibration and validation, to name a few as reported by Marelli et al. (See Figure 3)

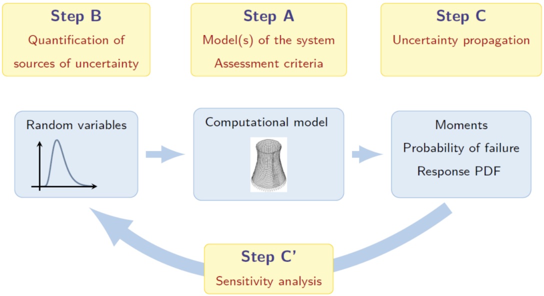

Figure 3: Sensitivity Analysis Pipelines in Marelli's paper

Sobol's method is based on decomposition of the model output variance into summands of variances of the input parameters in increasing dimensionality. As closely defined in Wikipedia: "From a black box perspective, any model may be viewed as a function $Y = f(X)$, where X is a vector of d uncertain model inputs $X1, X2, ...Xd$, and Y is a chosen univariate model output (note that this approach examines scalar model outputs, but multiple outputs can be analysed by multiple independent sensitivity analyses). Furthermore, it will be assumed that the inputs are independently and uniformly distributed within the unit hypercube."

$$Y = f_0 + \sum_{i=1}^{d} f_i\left(X_i\right) + \sum_{i<j}^{d} f_{ij}\left(X_i, X_j\right) + \cdots + f_{1,2,\ldots,d}\left(X_1, X_2, \ldots, X_d\right)$$

By proper factorization:

$$\operatorname{Var}(Y) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + \cdots + V_{12\ldots d}$$

$$\begin{aligned}
V_i &= \operatorname{Var}_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y \mid X_i\right)\right), \\
V_{ij} &= \operatorname{Var}_{X_{ij}}\left(E_{\mathbf{X}_{\cup ij}}\left(Y \mid X_i, X_j\right)\right) - \mathrm{V}_i - \mathrm{V}_j
\end{aligned}$$

We tried to build our own mathematical model to analysis the sensitivity of transformation from microscope to drill,
Translation part:

$$\varepsilon_{\vec{MD}} = \begin{bmatrix} I \\ -R_{MD}^{-1} R_X^{-1} sk(R_{TH}^{-1}(t_{TH} - t_{TD})) \\ -R_{MD}^{-1} sk(R_X^{-1} R_{TH}^{-1}(t_{TH} - t_{TD}) + R_X^{-1} t_X) \\ -R_{MD}^{-1} R_X^{-1} \\ -R_{MD}^{-1} \end{bmatrix}^T \begin{bmatrix} \varepsilon_{\vec{TD}} \\ \alpha_{\vec{TH}} \\ \alpha_{\vec{X}} \\ \varepsilon_{\vec{TH}} \\ \varepsilon_{\vec{X}} \end{bmatrix}$$

Rotation part:

$$\alpha_{\vec{MD}} = \left[I, \quad -R_{TD}^{-1} R_{TH} R_X^{-1}, \quad -R_{TD}^{-1} R_{TH}\right] \begin{bmatrix} \alpha_{\vec{TD}} \\ \alpha_{\vec{X}} \\ \alpha_{\vec{TH}} \end{bmatrix}$$

The formula is quite complex, and it requires cautious manual derivation. None of the package can do this automatically since the symbolic calculation needs to apply some special approximate rules as simplification.
While it‘s costly to build an accurate mathematical model, the UQLab gives us a method which can analyze a black-box system. All we need to do is to build a surrogate model such as PCE or Kriging with our own data. Using the ‘statistical inference’ module in UQLab to specify the probabilistic distribution of our inputs, then we can compute the Sobol‘s indices with the surrogate model we built at low cost to carry out sensitivity analysis.
There‘s still limited with the toolbox. To use the Sobol indices, input parameters should be statistically independent, while our input may not fulfill this requirement.

3 Reading List

1. L. C. French, M. S. Dietrich, and R. F. Labadie, “An estimate of the number of mastoidectomy procedures performed annually in the United States,” Ear Nose Throat J 87(5), 267–270 (2008).
2. J. Park, Q.-Y. Zhou, and V. Koltun, “Colored Point Cloud Registration Revisited,” in 2017 IEEE International Conference on Computer Vision (ICCV), pp. 143–152, IEEE, Venice (2017) [doi:10.1109/ICCV.2017.25].
3. F. Furrer et al., “Evaluation of Combined Time-Offset Estimation and Hand-Eye Calibration on Robotic Datasets,” in Field and Service Robotics, M. Hutter and R. Siegwart, Eds., pp. 145–159, Springer International Publishing, Cham (2018) [doi:10.1007/978-3-319-67361-5_10].
4. C. R. Razavi et al., “Image-Guided Mastoidectomy with a Cooperatively Controlled ENT Microsurgery Robot,” Otolaryngol Head Neck Surg 161(5), 852–855, SAGE Publications Inc (2019) [doi:10.1177/0194599819861526].
5. V. M. E. | 2841 N. H. R. Owensboro and K. 42303 | Office:691-6161, “Mastoid Surgery,” in Midwest Ear, Nose and Throat Head & Neck Surgery.
6. F. Dornaika and R. Horaud, “Simultaneous robot-world and hand-eye calibration,” IEEE Transactions on Robotics and Automation 14(4), 617–622 (1998) [doi:10.1109/70.704233].
7. S. Marelli and B. Sudret, “UQLab: A Framework for Uncertainty Quantification in Matlab,” 2554–2563, American Society of Civil Engineers (2014) [doi:10.1061/9780784413609.257].
8. A. Munawar et al., “Virtual Reality for Synergistic Surgical Training and Data Generation,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1–9 (2021) [doi:10.1080/21681163.2021.1999331].
9. Z. Chen, H. Liao, and X. Zhang, “Telecentric stereo micro-vision system: Calibration method and experiments,” Optics and Lasers in Engineering 57, 82–92 (2014) [doi:10.1016/j.optlaseng.2014.01.021].