Background Reading Report: Virtual reality for synergistic surgical training and data generation

Juan Antonio Barragan - jbarrag3 Hisashi Ishida - hishida3

April 5, 2022

1 Introduction

In this background reading report, the paper " *Virtual reality for synergistic surgical training and data generation*" will be summarized [4]. This paper was authored by Munawar, Adnan, and Li, Zhaoshuo and Kunjam, Punit and Nagururu, Nimesh and Ding, Andy S. and Kazanzides, Peter and Looi, Thomas and Creighton, Francis X. and Taylor, Russell H. and Unberath, Mathias and was published on November 26, 2021, in the Computer Methods in Biomechanics and Biomedical Engineering Journal. The following paper was selected because it introduces the skull surgery simulation platform in which our CIS project is currently being developed. In a few words, the described simulation allows training surgeons to perform a lateral skull-based surgical procedure by controlling a virtual drill. All the components of the simulation environment can be observed in figure 1.

Building on top of this simulation environment, we have proposed to introduce SDF-based (Sign distance function) feedback mechanisms to improve the situational awareness of the user while doing the training exercise. Specifically, we will be developing haptic and visual feedback mechanisms that will allow the user to understand how close he is to critical anatomical structures that should never be damaged with the drill. Understanding the paper reviewed in this report is a key step for the success of our project as it describes the software architecture as well as the advantages and limitations of the simulation environment. A good understanding of this information is key to providing a seamless integration of our SDF-based feedback mechanisms with the current simulation.

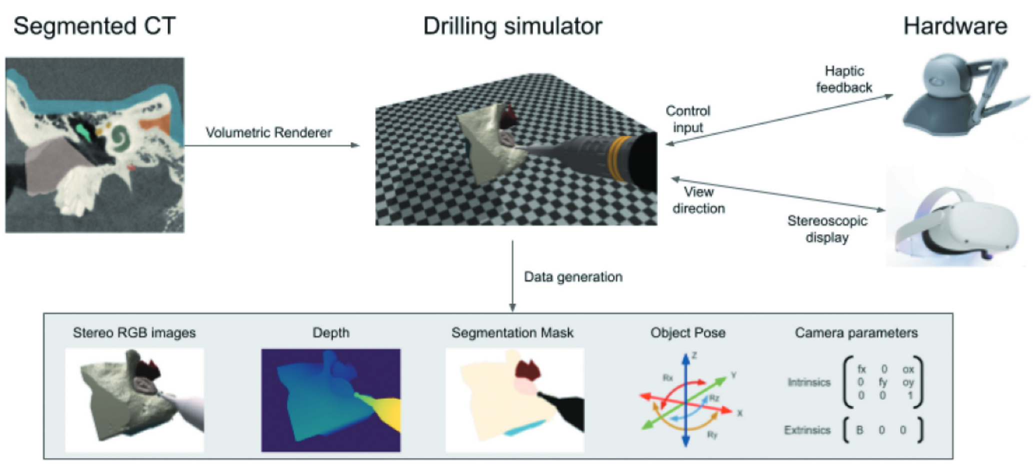

Figure 1: Simulation components

2 Technical summary of the paper

2.1 AMBF+/Simulation detail

The authors developed AMBF+ on top of AMBF and extended its functionality by (1) generating data (RGB stereo images, depth, and etc) in real-time and (2) proposing a plugin to customize the environment for different surgeries. Furthermore, for modular development, plugins are implemented and an extended rendering pipeline has enabled to generate and stream colored point-cloud and segmentation masks using custom-made OpenGL rendering.The framework utilizes an AMBF Description Format(ADF) to define rigid or soft bodies, sensors, actuators, joints and world scene objects. In order to adapt to the actual surgical scenario when surgical robots often have multiple interchangeable tools, ADF files are highly modular and a meta-data file, *launch file*, was used to load the models and world scene objects in AMBF+.

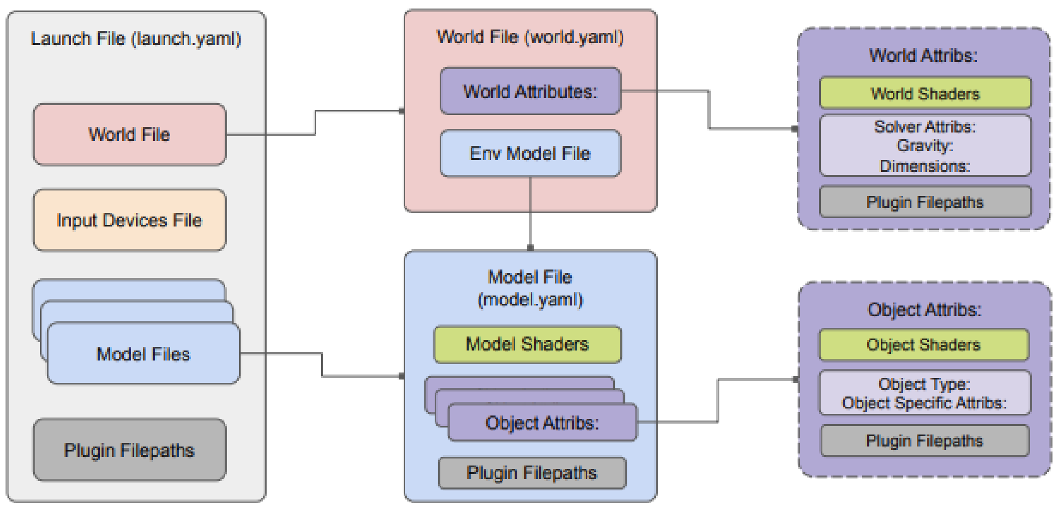

Figure 2: Flow of loading models and world scene objects into AMBF+

- Plugin Design

  AMBF+ is consisted of different hierarchical computational structures: *afSimulatorPlugin*, *afWorldPlugin*, *afModelPlugin*, and *afObjectPlugin*. Each of the plugins has a different interface but they are highly modulated and AMBF+ can simulate multi object scene in relatively effortless way.

- Rendering pipeline

  Previous AMBF uses OpenGL5 for rendering but could not take full advantage of some of its useful features such as loading custom shaders (vertex and fragment). AMBF+ utilizes OpenGL's Z-buffer and Framebuffer objects to generate the point cloud data and load special preprocessing shaders to assign a specific mask color to the visual object ignoring the lighting calculations.

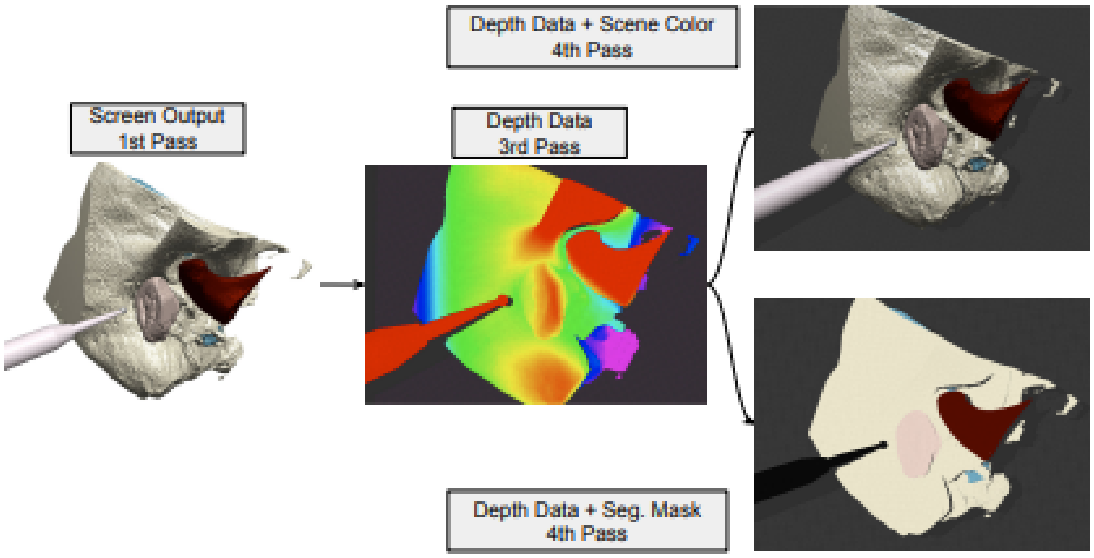

Figure 3: Point cloud representation and segmentation mask in AMBF+

- Data streaming

  They use Robot Operating System(ROS) interface and publish rostopic for the following data: Stereo image, depth, segmentation mask, object camera pose, camera parameter.

- Stereo display and VR support

  AMBF+ has support for Virtual Realty(VR) Head Mount Display(HBD) by generating a pair of stereoscopic images.

- Haptic device

  AMBF+ adopts CHAI3D's [2] finger proxy collision algorithm [5] to provide haptic feedback by simulating the collision with the surface of the volume. The movement of the tool will be simulated using the proxy/goal spheres such that the proxy sphere cannot cross the boundary of the surface or volume while the goal sphere will move freely to match with the tool. Force feedback will be generated proportional to the distance between the proxy and the goal spheres.

2.2 Simulation data evaluation

One of the biggest advantages of the AMBF+ simulation environment is that it can produce automatically annotated data for computer vision tasks while the user is performing the procedure. These annotated datasets can then be used to benchmark the performance of different algorithms in several computer vision paradigms. In particular, for this paper, two different tasks in computer vision are analyzed: (1) anatomy tracking and (2) stereo depth estimation.

2.2.1 Geometric-based computer vision algorithm anatomy tracking

The goal of anatomy tracking is to determine the pose of the anatomy with respect to the camera frame. This task has been of great interest in the augmented reality community as it would allow overlaying virtual objects on the world without requiring tracking markers. In this paper, the performance of the state-of-the-art ORB SLAM V3 algorithm [1] was evaluated using the data produced by the simulator. This algorithm was evaluated in two different settings: (1) static tool with camera motion, and (2) static camera with tool movements. The ORB SLAM V3 showed very robust tracking of the anatomy only when there was no camera motion. With a small motion of the camera, a tracking error of 40mm in translation and 8 degrees in rotation was observed. Moreover, it was observed that the algorithms failed completely in the tracking task when big camera motions were present. In situations, where the camera was kept fixed there was no degradation in the tracking algorithm due to the movement of the tools in the image.

2.2.2 Deep learning-based computer vision algorithm depth estimation

In this second application, the data from the simulator was used to train stereo depth estimation networks. These are deep learning models that can produce a depth map based on the pair of stereo images. The main advantage of using the AMBF+ for this task is that all the collected stereo images are accompanied by a corresponding ground-truth depth map. For this task the algorithm STTR [3] was trained with simulation data. The model after training produced a depth error of 1.98mm and was able to generalize to changes in the anatomy even when trained with a single anatomy. A sample prediction for this model can be observed in figure 4.

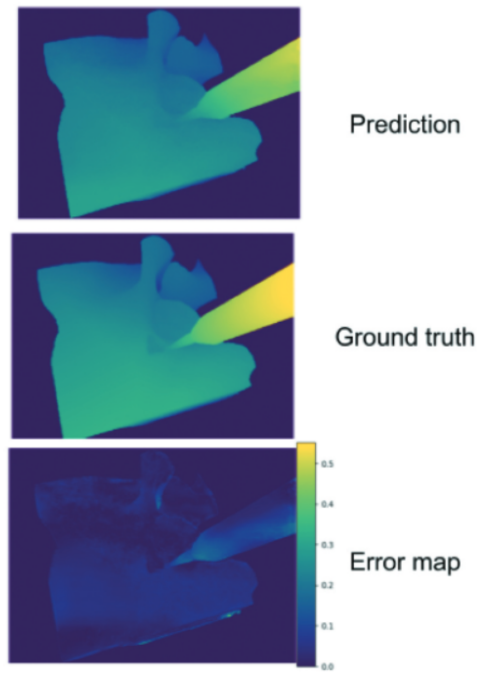

Figure 4: Depth prediction of the STTR model trained with the simulator data.

3 Discussion

In this report, we have illustrated that AMBF+ offers abundant real-time resources for downstream algorithm development and stereoscopic display on a virtual reality (VR) device and haptic feedback for immersive surgical simulation. However, the current simulation environment does not provide safety cues related to the distance between the drill and critical anatomies. Although a warning message is provided when the user collides with an anatomy, this is not enough feedback to teach the train how to avoid such a dangerous situation in a real procedure. Secondly, the simulator lacks the capability of providing haptic feedback to secure the patient safety. We believe that improving these two aspects by introducing a SDF based situational awareness methods will result in improved safety for the patient and reduce the workload of the surgeons.

References

[1] Carlos Campos, Richard Elvira, Juan J. Rodriguez, Jose M.M Montiel, and Juan D. Tardos. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

[2] Francois Conti, Federico Barbagli, Dan Morris, and Christopher Sewell. Chai 3d: An open-source library for the rapid development of haptic scenes. *IEEE World Haptics*, 38(1):21–29, 2005.

[3] Zhaoshuo Li, Xingtong Liu, Nathan Drenkow, Andy Ding, Francis X. Creighton, Russell H. Taylor, and Mathias Unberath. Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. *arXiv.org*, Aug 2021.

[4] Adnan Munawar, Zhaoshuo Li, Punit Kunjam, Nimesh Nagururu, Andy S. Ding, Peter Kazanzides, Thomas Looi, Francis X. Creighton, Russell H. Taylor, and Mathias Unberath. Virtual reality for synergistic surgical training and data generation. 0(0):1–9. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21681163.2021.1999331.

[5] Diego Ruspini and Oussama Khatib. Haptic display for human interaction with virtual dynamic environments. *Journal of Robotic Systems*, 18(12):769–783, 2001.