



Model-based 3D Tracking for Augmented Orthopedic Surgery

Albert Murienne¹, Boris Labbé¹ and Laurent Launay¹

¹IRT b-com, Rennes, France.
albert.murienne@b-com.com

Abstract

Current surgical navigation systems offer sub-millimetric real-time localization, however they are expensive, require the use of invasive markers attached to the patient, and often add extra operation time. In this paper we propose an affordable markerless navigation approach, based on mid end depth sensors, as an alternative to answer medical applications needs in terms of accuracy and robustness. An algorithm called Fast Volumetric Reconstruction (FaVoR) implements a compute-efficient approach for real time 3D model registration based tracking, allowing computed 3D poses to be used for video scene augmentation. After early testing with a first proof-of-concept implementation, a preliminary accuracy evaluation was performed using a dynamic test bench, achieving an average 2mm registration error during tracking.

1 Introduction

Augmented Reality (AR) provides an intuitive way to show clinical information to guide the surgeon during his procedure. For the additional information to be consistent with the scene, it has to be accurately positioned, and should be updated in real-time with the minimal possible latency (Navab, Traub, Sielhorst, Feuerstein, & Bichlmeier, 2007). Therefore, the key point is to have efficient object localization and tracking algorithm targeting, in the case of orthopedic surgery, the surgeon tools or rigid organs like bones.

For this study, a prototype application has been built to illustrate the scene augmentation needed to assist orthopedic knee cutting guide positioning. The algorithm performance is illustrated by displaying different type of hologram 3D shapes overlaid on the video stream: the planned knee alignment (Figure 1), the guide's drilling axis and the guide's cutting plane. The prototype relies on depth sensors, which have already shown good results on localization tasks (Ogor, et al., 2019). The application was first evaluated visually, then a test bench was designed to validate the approach and compare the results with traditional marker based navigation systems (Elfring, De La Fuente, & Radermacher, 2010).

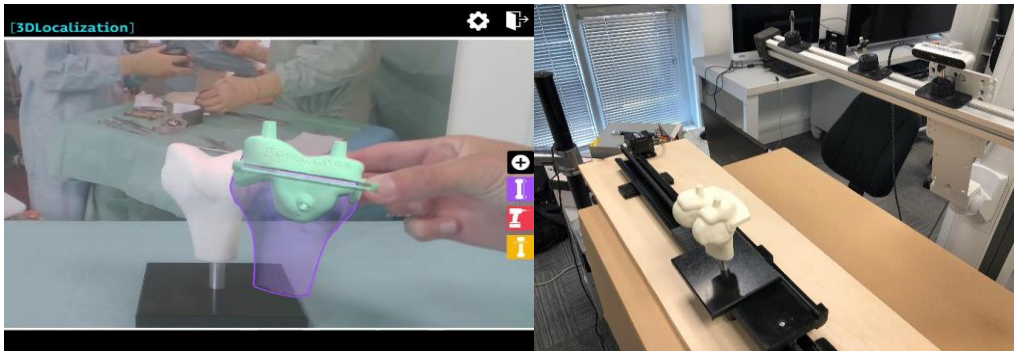


Figure 1: The prototype tracking application showing planned femur and PSI hologram, and the test bench.

2 Global Architecture

2.1 Software Architecture

FaVoR’s implementation is based on a complete geometrical paradigm, inspired by previous state-of-the-art real-time ICP tracking approaches like (Belshaw & Greenspan, 2009), (Korkalo & Kahn, 2016) and (Lee, et al., 2017), but using a ‘plane-to-plane’ Generalized ICP variant (Segal, Haehnel, & Thrun, 2009).

Our key performance indicator is the frame rate we can reach, which directly impacts the convergence speed and success of the ICP: the smaller the inter-frame displacement is, the faster the ICP will converge towards the right minima. The algorithm is implemented in C++, and thanks to heavy use of vectorization and parallelization, it can reach up to 60 frames-per-second with an input camera depth resolution of 640x480 (~300K pixels). The tracked objects are typically hand manipulated. So considering an average hand speed of 1m/s during movement, at 60fps, the handheld object’s inter-frame displacement is about 16mm, which gives us a good idea of the search area required to support this speed. This search area is configurable using a dedicated parameter, the search radius SR, which influence is further described in this paper.

Another problem is that raw computed poses can suffer from detection noise coming from the depth sensor, so exponential smoothing was applied for more realistic and less shaky results (LaViola, 2003), as well as limited vacillation when the object remains steady. We didn’t investigate outlier detection based filtering yet, as exponential smoothing showed some satisfying results for now.

2.2 Hardware Architecture

A prototype application was developed using an Intel RealSense D415 depth camera, running on a Microsoft Surface Pro 5 with core i7 processor. The algorithm is CPU only (no GPU processing), therefore it does not require any external processing server. The camera is attached to the back of the tablet, allowing the application to run in “video see-through” augmentation mode.

The goal of our test bench is to have reproducible and configurable movement (amplitude and speed), in order to ensure that our approach is robust to fast movement. The current version of the bench

can reach up to 0.15m/s speed translation, with the help of a stepping motor controlled cart moving along a rail, as seen in Figure 1.

3 Results

A first study has been carried out, targeting registration residual error (mean error over matched points), during static or dynamic translations phases, at different speeds and different algorithmic configurations, and tracking a prototype femoral PSI (~8cm span) at 40cm distance. The influence of search radius (SR) parameter was evaluated, using an Intel Realsense D415 and an Occipital Structure Core running in 640x480@60fps depth mode. During dynamic runs, a 20cm back and forth translation was performed with configurable speed. The results are illustrated in Figure 2, where the RMSE registration error is reported for each test configuration. Blank cells indicate tracking was lost during test.

<i>Intel Realsense D415</i>	0 cm/s	5 cm/s	10 cm/s	15 cm/s
SR=3mm	1.68mm ±0.02mm	1.71mm ±0.03mm	-	-
SR=5mm	2.12mm ±0.03mm	2.16mm ±0.06mm	2.23mm ±0.14mm	-
SR=8mm	2.83mm ±0.08mm	3.02mm ±0.16mm	3.14mm ±0.18mm	3.16mm ±0.19mm
<i>Occipital Structure Core</i>	0 cm/s	5 cm/s	10 cm/s	15 cm/s
SR=3mm	1.59mm ±0.01mm	1.60mm ±0.03mm	-	-
SR=5mm	1.92mm ±0.04mm	1.87mm ±0.07mm	1.93mm ±0.085mm	2.05mm ±0.15mm
SR=8mm	2.34mm ±0.09mm	2.46mm ±0.15mm	2.55mm ±0.165mm	2.75mm ±0.295mm

Figure 2 : Influence of FaVoR search radius on tracking robustness and accuracy.

These results show that search radius has direct influence on the tradeoff between robustness and accuracy: high search radius means less accuracy due to more potential outliers included in the search path, small search radius means less robustness to fast inter-frame movement. The direct correlation between search radius and speed is not a straight forward computation: 15cm/s should be supported with SR 3mm, but that would be neglecting sensors inherent noise and lag. When comparing the two cameras, the Structure Core is overall slightly more accurate (up to 0.5mm at highest speed) and more robust (SR 5mm@15cm/s didn't lose tracking). The Structure Core has a global shutter, whereas the D415 has a rolling shutter, so it's more sensitive to moving scenes. The overall mean accuracy level is around 2-3mm. Further work should assess the orientation error by collecting ground truth data.

4 Conclusion

In this paper, a fast volumetric registration technique embedded in a depth sensor equipped tablet has been presented and evaluated on a dynamic test bench. The system can reach 60fps on the tablet CPU and a tracking registration accuracy of 2mm. Robustness improvement is still needed to support higher speed movements. Future work will include ground truth based accuracy qualification and

further code optimizations to address cameras 90fps mode. Machine learning based pose inference (Peng, Liu, Huang, Bao, & Zhou, 2018) could also be investigated to initialize registration.

Acknowledgment

This work benefited from State aid managed by the National Research Agency under the future investment program bearing the reference ANR-17-RHUS-0005 (FollowKnee project).

References

- Belshaw, M., & Greenspan, M. (2009). A high speed iterative closest point tracker on an FPGA platform. *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, p.1449-1456.
- Elfring, R., De La Fuente, M., & Radermacher, K. (2010). Assessment of optical localizer accuracy for computer aided surgery systems. *Computer Aided Surgery*, Volume 15, Issue 1-3, p.1-12.
- Korkalo, O., & Kahn, S. (2016). Real-time depth camera tracking with CAD models and ICP. *Journal of Virtual Reality and Broadcasting*.
- LaViola, J. (2003). Double Exponential Smoothing: An Alternative to Kalman Filter-Based Predictive Tracking. *Eurographics Workshop on Virtual Environments*, p.199-206.
- Lee, S. C., Fuerst, B., Tateno, K., Johnson, A., Fotouhi, J., Osgood, G., . . . Navab, N. (2017). Multi-modal imaging, model-based tracking, and mixed reality visualisation for orthopaedic surgery. *Healthcare technology letters 4*, p.168-173.
- Navab, N., Traub, J., Sielhorst, T., Feuerstein, M., & Bichlmeier, C. (2007). Action and Workflow-Driven Augmented Reality for Computer-Aided Medical Procedures. *IEEE computer graphics and applications*, 27. 10-4. 10.1109/MCG.2007.117.
- Ogor, J., Dardenne, G., Sta, S., Bert, J., Letissier, H., Stindel, E., & Hamitouche, C. (2019). 3D Pose Estimation with Depth Camera for Markerless Computer Assisted Orthopaedic Surgery. *CAOS 19th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery*, vol 3, p.289-292.
- Peng, S., Liu, Y., Huang, Q., Bao, H., & Zhou, X. (2018). Pvnet: Pixel-wise voting network for 6dof pose estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Segal, A., Haehnel, D., & Thrun, S. (2009). Generalized-ICP. *Proceedings of Robotics: Science and Systems*, 10.15607/RSS.2009.V.021.