Markerless Inside-Out Tracking for 3D Ultrasound Compounding

Benjamin Busam[1,2(✉)], Patrick Ruhkamp[1,2], Salvatore Virga[1],
Beatrice Lentes[1], Julia Rackerseder[1], Nassir Navab[1,3],
and Christoph Hennersperger[1]

[1] Computer Aided Medical Procedures, Technische Universität München,
Munich, Germany
`salvo.virga@tum.de`
[2] FRAMOS GmbH, Taufkirchen, Germany
`{b.busam,p.ruhkamp}@framos.com`
[3] Computer Aided Medical Procedures, Johns Hopkins University,
Baltimore, USA

Abstract. Tracking of rotation and translation of medical instruments plays a substantial role in many modern interventions and is essential for 3D ultrasound compounding. Traditional external optical tracking systems are often subject to line-of-sight issues, in particular when the region of interest is difficult to access. The introduction of inside-out tracking systems aims to overcome these issues. We propose a marker-less tracking system based on visual SLAM to enable tracking of ultrasound probes in an interventional scenario. To achieve this goal, we mount a miniature multi-modal (mono, stereo, active depth) vision system on the object of interest and relocalize its pose within an adaptive map of the operating room. We compare state-of-the-art algorithmic pipelines and apply the idea to transrectal 3D ultrasound (TRUS). Obtained volumes are compared to reconstruction using a commercial optical tracking system as well as a robotic manipulator. Feature-based binocular SLAM is identified as the most promising method and is tested extensively in challenging clinical environments and for the use case of prostate US biopsies.

Keywords: 3D ultrasound imaging · Line-of-sight avoidance
Visual inside-out tracking · SLAM · Computer assisted interventions

1 Introduction

Tracking of medical instruments and tools is required for various systems in medical imaging, as well as computer aided interventions. Especially for medical applications such as 3D ultrasound compounding, accurate tracking is an important requirement, however often comes with severe drawbacks impacting the medical workflow. Mechanical tracking systems can provide highly precise
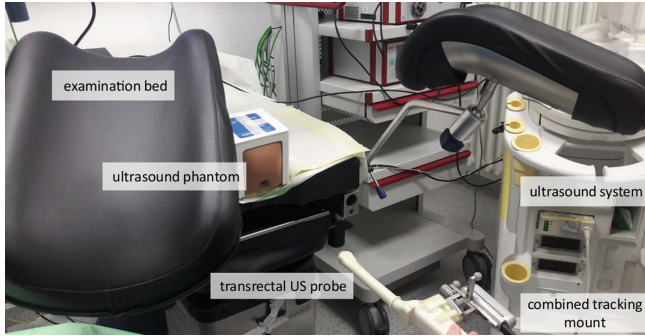
Fig. 1. Interventional setup for fusion biopsy. Clinical settings are often characterized by cluttered setups with tools and equipment around the examination bed. While such environments are challenging for outside-in tracking, they can provide a rich set of features for SLAM-based inside-out tracking.

tracking through a kinematic chain [1]. These systems often require bulky and expensive equipment, which cannot be adapted to a clinical environment where high flexibility needs to be ensured. In contrast to that, electromagnetic tracking is flexible in its use, but limited to comparably small work spaces and can interfere with metallic objects in proximity to the target, reducing the accuracy [2].

Optical tracking systems (OTS) enjoy widespread use as they do not have these disadvantages. Despite favourable spatial accuracy under optimal conditions, respective systems suffer from constraints by the required line-of-sight. Robust marker based methods such as [3] address this problem and work even if the target is only partly visible. However, the marker-visibility issue is further complicated for imaging solutions relying on tracking systems, with prominent examples being freehand SPECT [4] as well as freehand 3D ultrasound [5].

Aiming at both accurate and flexible systems for 3D imaging, a series of developments have been proposed recently. Inside-out tracking for collaborative robotic imaging [6] proposes a marker-based approach using infrared cameras, however, not resolving line-of-sight issues. A first attempt at making use of localized features employs tracking of specific skin features for estimation of 3D poses [7] in 3D US imaging. While this work shows promising results, it is constrained to the specific anatomy at hand.

In contrast to previous works, our aim is to provide a generalizable tracking approach without requiring a predefined or application-specific set of features while being easy to setup even for novice users. With the recent advent of advanced miniaturized camera systems, we evaluate an inside-out tracking approach solely relying on features extracted from image data for pose tracking (Fig. 1).

For this purpose, we propose the use of visual methods for simultaneously mapping the scenery and localizing the system within it. This is enabled by building up a map from characteristic structures within the previously unknown scene observed by a camera, which is known as SLAM [8]. Different image
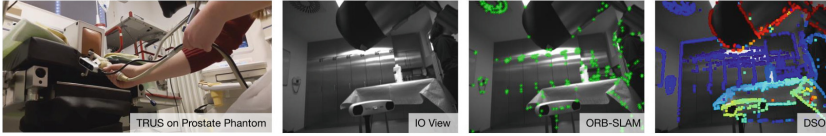
Fig. 2. 3D TRUS volume acquisition of prostate phantom. An inside-out camera is mounted on a transrectal US transducer together with a rigid marker for an outside-in system in the prostate biopsy OR. The consecutive images show the relevant extracted data for the considered SLAM methods.

modalities can be used for visual SLAM and binocular stereo possesses many benefits compared to monocular vision or active depth sensors.

On this foundation, we propose a flexible inside-out tracking approach relying on image features and poses retrieved from SLAM. We evaluate different methods in direct comparison to a commercial tracking solution and ground truth, and show an integration for freehand 3D US imaging as one potential use-case. The proposed prototype is the first proof of concept for SLAM-based inside-out tracking for interventional applications, applied here to 3D TRUS as shown in Fig. 2. The novelty of pointing the camera away from the patient into the quasi-static room while constantly updating the OR map enables advantages in terms of robustness, rotational accuracy and line-of-sight problem avoidance. Thus, no hardware relocalization of external outside-in systems is needed, partial occlusion is handled with wide-angle lenses and the method copes with dynamic environmental changes. Moreover, it paves the path for automatic multi-sensor alignment through a shared common map while maintaining an easy installation by clipping the sensor to tools.

2   Methods

For interventional imaging and specifically for the case of 3D ultrasound, the goal is to provide rigid body transformations of a desired target with respect to a common reference frame. This way, we denote ${}^{B}\mathbf{T}_{A}$ as transformation A to B. On this foundation, the transformation ${}^{W}\mathbf{T}_{US}$ from the ultrasound image (US) should be indicated in a desired world coordinate frame (W). For the case of inside-out based tracking - and in contrast to outside-in approaches - the ultrasound probe is rigidly attached to the camera system, providing the desired relation to the world reference frame

$$ {}^{W}\mathbf{T}_{US} = {}^{W}\mathbf{T}_{RGB} \cdot {}^{RGB}\mathbf{T}_{US}, \tag{1} $$

where ${}^{W}\mathbf{T}_{RGB}$ is retrieved from tracking. The static transformation ${}^{RGB}\mathbf{T}_{US}$ can be obtained with a conventional 3D US calibration method [9].

Inside-out tracking is proposed on the foundation of a miniature camera setup as described in Sect. 3. The setup provides different image modalities for the visual SLAM. Monocular SLAM is not suitable for our needs, since it

needs an appropriate translation without rotation within the first frames for proper initialization and suffers from drift due to accumulating errors over time. Furthermore, the absolute scale of the reconstructed map and the trajectory is unknown due to the arbitrary baseline induced by the non-deterministic initialization for finding a suitable translation. Relying on the depth data from the sensor would not be sufficient for the desired tracking accuracy, due to noisy depth information. A stereo setup can account for absolute scale by a known fixed baseline and movements with rotations only can be accounted for since matched feature points can be triangulated for each frame.

For the evaluations we run experiments with publicly available SLAM methods for better reproducibility and comparability. ORB-SLAM2 [8] is used as state-of-the-art feature based method. The well-known direct methods [10,11] are not eligible due to the restriction to monocular cameras. We rely on the recent publicly available[1] stereo implementation of Direct Sparse Odometry (DSO) [12].

The evaluation is performed with the coordinate frames depicted in Fig. 3. The intrinsic camera parameters of the involved monocular and stereo cameras (RGB, IR1, IR2) are estimated as proposed by [13]. For the rigid transformation from the robotic end effector to the inside-out camera, we use the hand-eye calibration algorithm of Tsai-Lenz [14] in eye-on-hand variant implemented in ViSP [15] and the eye-on-base version to obtain the rigid transformation from the optical tracking system to the robot base. To calibrate the ultrasound image plane with respect to the different tracking systems, we use the open source PLUS ultrasound toolkit [16] and provide a series of correspondence pairs using a tracked stylus pointer.

3   Experiments and Validation

To validate the proposed tracking approach, we first evaluate the tracking accuracy, followed by a specific analysis for the suitability to 3D ultrasound imaging. We use a KUKA iiwa (KUKA Roboter GmbH, Augsburg, Germany) 7 DoF robotic arm to gather ground truth tracking data which guarantees a positional reproducibility of ±0.1 mm. To provide a realistic evaluation, we also utilize an optical infrared-based outside-in tracking system (Polaris Vicra, Northern Digital Inc., Waterloo, Canada). Inside-out tracking is performed with the Intel RealSense Depth Camera D435 (Mountain View, US), providing RGB and infrared stereo data in a portable system (see Fig. 3). Direct and feature based SLAM methods for markerless inside-out tracking are compared and evaluated against marker based optical inside-out tracking with ArUco [17] markers (16×16 cm) and classical optical outside-in tracking. For a quantitative analysis, a combined marker with an optical target and a miniature vision sensor is attached to the robot end effector. The robot is controlled using the Robot Operating System (ROS) while the camera acquisition is done on a separate machine

---

[1] https://github.com/JiatianWu/stereo-dso, Horizon Robotics, Inc. Beijing, China, Authors: Wu, Jiatian; Yang, Degang; Yan, Qinrui; Li, Shixin.
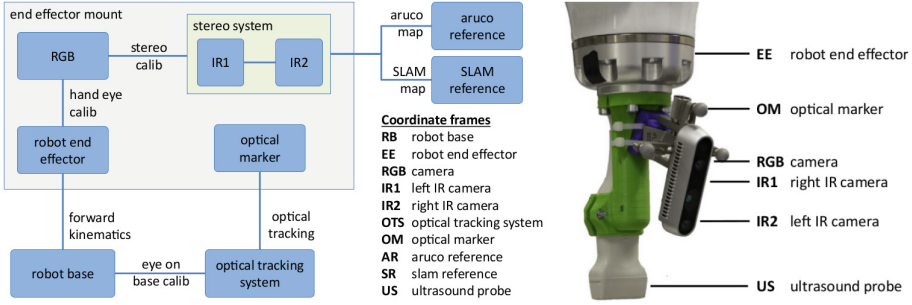
Fig. 3. System architecture and coordinate frames. Shown are all involved coordinate reference frames to evaluate the system performance (**left**) as well as the specific ultrasound mount used for validation, integrating optical and camera-based tracking with one attachable target (**right**).
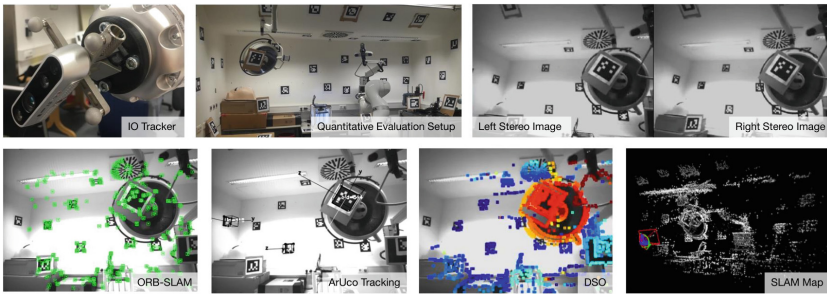

Fig. 4. Quantitative evaluation setup. The first row illustrates the operating room where the quantitative analysis is performed together with the inside-out stereo view. The second row depicts various calculated SLAM information necessary to create the map.

using the intel RealSense SDK[2]. The pose of the RGB camera and the tracking target are communicated via TCP/IP with a publicly available library[3]. The images are processed on an intel Core i7-6700 CPU, 64bit, 8 GB RAM running Ubuntu 14.04. We use the same constraints as in a conventional TRUS. Thus, the scanning time, covered volume and distance of the tracker is directly comparable and the error analysis reflects this specific procedure with all involved components. Figure 4 shows the clinical environment for the quantitative evaluation together with the inside-out view and the extracted image information for the different SLAM methods.

3.1 Tracking Accuracy

To evaluate the tracking accuracy, we use the setup described above and acquire a series of pose sequences. The robot is programmed to run in gravity compensation mode such that it can be directly manipulated by a human operator. The forward kinematics of a robotic manipulator are used as ground truth (GT) for the actual movement.

To allow for error evaluation, we transform all poses of the different tracking systems in the joint coordinate frame coinciding at the RGB-camera of the end effector mount (see Fig. 3 for an overview of all reference frames)

$$^{RGB}\mathbf{T}_{RB} = {}^{RGB}\mathbf{T}_{EE} \cdot {}^{EE}\mathbf{T}_{RB} \tag{2}$$

$$^{RGB}\mathbf{T}_{SR} = {}^{RGB}\mathbf{T}_{EE} \cdot {}^{EE}\mathbf{T}_{RB} \cdot {}^{RB}\mathbf{T}_{IR1,0} \cdot {}^{IR1,0}\mathbf{T}_{SR} \tag{3}$$

$$^{RGB}\mathbf{T}_{AR} = {}^{RGB}\mathbf{T}_{EE} \cdot {}^{EE}\mathbf{T}_{RB} \cdot {}^{RB}\mathbf{T}_{IR1,0} \cdot {}^{IR1,0}\mathbf{T}_{AR} \tag{4}$$

$$^{RGB}\mathbf{T}_{OTS} = {}^{RGB}\mathbf{T}_{EE} \cdot {}^{EE}\mathbf{T}_{RB} \cdot {}^{RB}\mathbf{T}_{OTS} \cdot {}^{OTS}\mathbf{T}_{OM}, \tag{5}$$

providing a direct way to compare the optical tracking system (OTS), to SLAM-based methods (SR), and the ArUco-based tracking (AR).

In overall, 5 sequences were acquired with a total of 8698 poses. The pose error for all compared system is indicated in Fig. 5, where the translation error is given by the RMS of the residuals compared with the robotic ground truth while the illustrated angle error gives angular deviation of the rotation axis. From the results it can be observed that optical tracking provides the best results, with translation errors of 1.90 ± 0.53 mm, followed by 2.65 ± 0.74 mm for ORB-SLAM and 3.20 ± 0.96 for DSO, ArUco with 5.73 ± 1.44 mm. Interestingly, the SLAM-based methods provide better results compared to OTS, with errors of $1.99 \pm 1.99°$ for ORB-SLAM, followed by $3.99 \pm 3.99°$ for DSO, respectively. OTS estimates result in errors of $8.43 \pm 6.35°$, and ArUco orientations are rather noisy with $29.75 \pm 48.92°$.
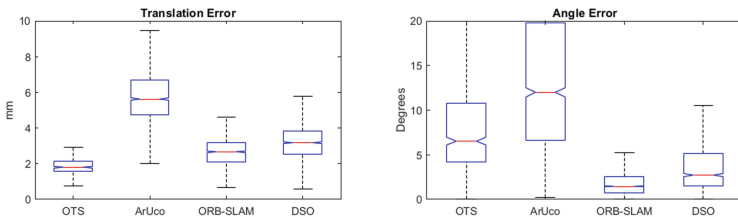

Fig. 5. Comparison of tracking error. Shown are translational and rotational errors compared to ground truth for all evaluated systems.

3.2   Markerless Inside-Out 3D Ultrasound

On the foundation of favourable tracking characteristics, we evaluate the performance of a markerless inside-out 3D ultrasound system by means of image quality and reconstruction accuracy for a 3D US compounding. For imaging, the tracking mount shown in Fig. 3 is integrated with a 128 elements linear transducer (CPLA12875, 7 MHz) connected to a cQuest Cicada scanner (Cephasonics, CA, USA). For data acquisition, a publicly available real-time framework is employed.[4] We perform a sweep acquisition, comparing OTS outside-in tracking with the proposed inside-out method and evaluate the quality of the reconstructed data while we deploy [18] for temporal pose synchronization. Figure 6 shows a qualitative comparison of the 3D US compoundings for the same sweep with the different tracking methods.[5]
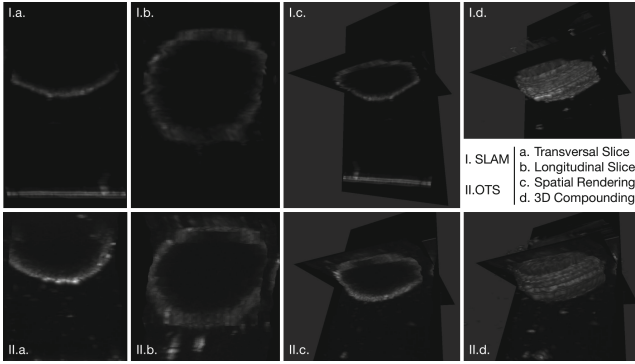

Fig. 6. Visualization of 3D US compounding quality. Shown are longitudinal and transversal slices as well as a 3D rendering of the resulting reconstructed 3D data from a tracked ultrasound acquisition of a ball phantom for the proposed tracking using ORB-SLAM in comparison with a commercial outside-in OTS. The structure appears spherically while the rotational accuracy advantage of ORB-SLAM causes a smoother rendering surface and a more clearly defined phantom boundary in the computed slices.

4   Discussion and Conclusion

From our evaluation, it appears that ArUco markers are viable only for approximate positioning within a room rather than accurate tracking. Our proposed inside-out approach shows valuable results compared to standard OTS and even outperforms the outside-in system in terms of rotational accuracy. These findings concur with assumptions based on the camera system design, as small rotations close to the optical principal point of the camera around any axis will lead to

---

[4] https://github.com/IFL-CAMP/supra.

[5] A video analysis of the method can be found here: https://youtu.be/SPy5860K49Q.

severe changes in the viewing angle, which can visually be described as inside-out rotation leverage effect.

One main advantage of the proposed methods is with respect to usability in practice. By not relying on specific markers, there is no need for setting up an external system or a change in setup during procedures. Additionally, we can avoid line-of-sight problems, and potentially allow for highly accurate tracking even for complete rotations around the camera axis without loosing tracking. This is in particular interesting for applications that include primarily rotation such as transrectal prostate fusion biopsy. Besides the results above, our proposed method is capable of orientating itself within an unknown environment by mapping its surrounding from the beginning of the procedure. This mapping is build up from scratch without the necessity of any additional calibration. Our tracking results for a single sensor also suggest further investigation towards collaborative inside-out tracking with multiple systems at the same time, orientating themselves within a global map as common reference frame.

References

1. Hennersperger, C., et al.: Towards MRIs-based autonomous robotic US acquisitions: a first feasibility study. MI **36**(2), 538–548 (2017)
2. Kral, F., Puschban, E.J., Riechelmann, H., Freysinger, W.: Comparison of optical and electromagnetic tracking for navigated lateral skull base surgery. IJMRCAS **9**(2), 247–252 (2013)
3. Busam, B., Esposito, M., Che'Rose, S., Navab, N., Frisch, B.: A stereo vision approach for cooperative robotic movement therapy. In: ICCVW, pp. 127–135 (2015)
4. Heuveling, D., Karagozoglu, K., Van Schie, A., Van Weert, S., Van Lingen, A., De Bree, R.: Sentinel node biopsy using 3D lymphatic mapping by freehand spect in early stage oral cancer: a new technique. CO **37**(1), 89–90 (2012)
5. Fenster, A., Downey, D.B., Cardinal, H.N.: Three-dimensional ultrasound imaging. Phys. Med. Biol. **46**(5), R67 (2001)
6. Esposito, M., et al.: Cooperative robotic gamma imaging: enhancing US-guided needle biopsy. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 611–618. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_73
7. Sun, S.-Y., Gilbertson, M., Anthony, B.W.: Probe localization for freehand 3D ultrasound by tracking skin features. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 365–372. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10470-6_46
8. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an open-source slam system for monocular, stereo, and RGB-D cameras. TR **33**(5), 1255–1262 (2017)
9. Hsu, P.W., Prager, R.W., Gee, A.H., Treece, G.M.: Freehand 3D ultrasound calibration: a review. In: Sensen, C.W., Hallgrímsson, B. (eds.) Advanced Imaging in Biology and Medicine, pp. 47–84. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-68993-5_3
10. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 834–849. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_54

11. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. PAMI (2018)
12. Wang, R., Schwörer, M., Cremers, D.: Stereo DSO: large-scale direct sparse visual odometry with stereo cameras. In: ICCV (2017)
13. Zhang, Z.: A flexible new technique for camera calibration. PAMI **22**(11), 1330–1334 (2000)
14. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. TRA **5**(3), 345–358 (1989)
15. Marchand, É., Spindler, F., Chaumette, F.: Visp for visual servoing: a generic software platform with a wide class of robot control skills. RAM **12**(4), 40–52 (2005)
16. Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., Fichtinger, G.: Plus: open-source toolkit for ultrasound-guided intervention systems. BE **61**(10), 2527–2537 (2014)
17. Garrido-Jurado, S., noz Salinas, R.M., Madrid-Cuevas, F., Marín-Jiménez, M.: Automatic generation and detection of highly reliable fiducial markers under occlusion. PR **47**(6), 2280–2292 (2014)
18. Busam, B., Esposito, M., Frisch, B., Navab, N.: Quaternionic upsampling: Hyperspherical techniques for 6 DoF pose tracking. In: 3DV, IEEE (2016) 629–638