Closing the Calibration Loop: An Inside-Out-Tracking Paradigm for Augmented Reality in Orthopedic Surgery

Jonas Hajek[1,2], Mathias Unberath[1(✉)], Javad Fotouhi[1], Bastian Bier[1,2],
Sing Chun Lee[1], Greg Osgood[3], Andreas Maier[2], Mehran Armand[4],
and Nassir Navab[1]

[1] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, USA
unberath@jhu.edu
[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg,
Erlangen, Germany
[3] Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, USA
[4] Applied Physics Laboratory, Johns Hopkins University, Baltimore, USA

Abstract. In percutaneous orthopedic interventions the surgeon attempts to reduce and fixate fractures in bony structures. The complexity of these interventions arises when the surgeon performs the challenging task of navigating surgical tools percutaneously only under the guidance of 2D interventional X-ray imaging. Moreover, the intra-operatively acquired data is only visualized indirectly on external displays. In this work, we propose a flexible Augmented Reality (AR) paradigm using optical see-through head mounted displays. The key technical contribution of this work includes the marker-less and dynamic tracking concept which closes the calibration loop between patient, C-arm and the surgeon. This calibration is enabled using Simultaneous Localization and Mapping of the environment, i.e. the operating theater. In return, the proposed solution provides *in situ* visualization of pre- and intra-operative 3D medical data directly at the surgical site. We demonstrate pre-clinical evaluation of a prototype system, and report errors for calibration and target registration. Finally, we demonstrate the usefulness of the proposed inside-out tracking system in achieving "bull's eye" view for C-arm-guided punctures. This AR solution provides an intuitive visualization of the anatomy and can simplify the hand-eye coordination for the orthopedic surgeon.

Keywords: Augmented Reality · Human computer interface
Intra-operative visualization and guidance · C-arm · Cone-beam CT

J. Hajek, M. Unberath and J. Fotouhi—These authors are considered joint first authors.

1   Introduction

Modern orthopedic trauma surgery focuses on percutaneous alternatives to many complicated procedures [1,2]. These minimally invasive approaches are guided by intra-operative X-ray images that are acquired using mobile, non-robotic C-arm systems. It is well known that X-ray images from multiple orientations are required to warrant understanding of the 3D spatial relations since 2D fluoroscopy suffers from the effects of projective transformation. Mastering the mental mapping of tools to anatomy from 2D images is a key competence that surgeons acquire through extensive training. Yet, this task often challenges even experienced surgeons leading to longer procedure times, increased radiation dose, multiple tool insertions, and surgeon frustration [3,4].

If 3D pre- or intra-operative imaging is available, challenges due to indirect visualization can be mitigated, substantially reducing surgeon task load and fostering improved surgical outcome. Unfortunately, most of the previously proposed systems provide 3D information at the cost of integrating outside-in tracking solutions that require additional markers and intra-operative calibration that hinder clinical acceptance [3]. As an alternative, intuitive and real-time visualization of 3D data in Augmented Reality (AR) environments has recently received considerable attention [4,5]. In this work, we present a purely image-based inside-out tracking concept and prototype system that dynamically closes the calibration loop between surgeon, patient, and C-arm enabling intra-operative optical see-through head-mounted display (OST HMD)-based AR visualization overlaid with the anatomy of interest. Such *in situ* visualization could benefit residents in training that observe surgery to fully understand the actions of the lead surgeon with respect to the deep-seated anatomical targets. These applications in addition to simple task such as optimal positioning of C-arm systems, do not require the accuracy needed for surgical navigation and, therefore, could be the first target for OST HMD visualization in surgery. To the best of our knowledge, this prototype constitutes the first marker-less solution to intra-operative 3D AR on the target anatomy.

2   Materials and Methods

2.1   Calibration

The inside-out tracking paradigm, core of the proposed concept, is driven by the observation that all relevant entities (surgeon, patient, and C-arm) are positioned relative to the same environment, which we will refer to as the "world coordinate system". For intra-operative visualization of 3D volumes overlaid with the patient, we seek to dynamically recover

$$^{S}\mathbf{T}_{V}(t) = {}^{S}\mathbf{T}_{W} \underbrace{\left({}^{T}\mathbf{T}_{W}^{-1}(t_0) \; {}^{T}\mathbf{T}_{C}(t_0)\right) \; {}^{V}\mathbf{T}_{C}^{-1}}_{^{W}\mathbf{T}_{V}}, \tag{1}$$

the transformation describing the mapping from the surgeon's eyes to the 3D image volume. In Eq. 1, t_0 describes the time of pre- to intra-operative image

Fig. 1. Spatial relations that must be estimated dynamically to enable the proposed AR environment. Transformations shown in black are estimated directly while transformations shown in orange are derived.

registration while t is the current time point. The spatial relations that are required to dynamically estimate $^{\mathrm{S}}\mathbf{T}_{\mathrm{V}}$ are explained in the remainder of this section and visualized in Fig. 1.

$^{W}\mathbf{T}_{S/T}$: The transformations $^{W}\mathbf{T}_{\mathrm{S/T}}$ are estimated using Simultaneous Localization and Mapping (SLAM) thereby incrementally constructing a map of the environment, i.e. the world coordinate system [6]. Exemplarily for the surgeon, SLAM solves

$$^{\mathrm{W}}\mathbf{T}_{\mathrm{S}}(t) = \underset{\mathrm{W}\hat{\mathbf{T}}_{\mathrm{S}}}{\arg\min} \; \mathrm{d}\Big(\mathbf{f}_{\mathrm{W}}\left(\mathbf{P}\,^{\mathrm{W}}\hat{\mathbf{T}}_{\mathrm{S}}(t)\mathbf{x}_{\mathrm{S}}(t)\right), \mathbf{f}_{\mathrm{S}}(t)\Big), \qquad (2)$$

where $\mathbf{f}_{\mathrm{S}}(t)$ are features in the image at time t, $\mathbf{x}_{\mathrm{S}}(t)$ are the 3D locations of these feature estimates either via depth sensors or stereo, $\mathbf{P}$ is the projection operator, and $d(\cdot,\cdot)$ is the feature similarity to be optimized. A key innovation of this work is the inside-out SLAM-based tracking of the C-arm w.r.t. the *exact same* map of the environment by means of an additional tracker rigidly attached to the C-shaped gantry. This becomes possible if both trackers are operated in a master-slave configuration and observe partially overlapping parts of the environment, i.e. a feature rich and temporally stable area of the environment. This suggests, that the cameras on the C-arm tracker (in contrast to previous solutions [5,7]) need to face the room rather than the patient.

$^{T}\mathbf{T}_{C}$: The tracker is rigidly mounted on the C-arm gantry suggesting that one-time offline calibration is possible. Since the X-ray and tracker cameras have no overlap, methods based on multi-modal patterns as in [4,5,7] fail. However, if

poses of both cameras w.r.t. the environment and the imaging volume, respectively, are known or can be estimated, Hand-Eye calibration is feasible [8]. Put concisely, we estimate a rigid transform $^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ such that $\mathbf{A}(t_i)\,^{\mathrm{T}}\mathbf{T}_{\mathrm{C}} = \,^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}\mathbf{B}(t_i)$, where $(\mathbf{A}/\mathbf{B})(t_i)$ is the relative pose between subsequent poses at times $i, i{+}1$ of the tracker and the C-arm, respectively. Poses of the C-arm $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}(t_i)$ are known because our prototype (Sect. 2.2) uses a cone-beam CT (CBCT) enabled C-arm with pre-calibrated circular source trajectory such that several poses $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ are known. During one sweep, we estimate the poses of the tracker $^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_i)$ via Eq. 1. Finally, we recover $^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$, and thus $^{\mathrm{W}}\mathbf{T}_{\mathrm{C}}$, as detailed in [8].

$^{\mathrm{V}}\mathbf{T}_{C}$: To close the loop by calibrating the patient to the environment, we need to estimate the $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ describing the transformation from 3D image volumes to an intra-operatively acquired X-ray image. For pre-operative data, $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ can be estimated via image-based 3D/2D registration, e.g. as in [9,10]. If the C-arm is CBCT capable and the 3D volume is acquired intra-procedurally, $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ is known and can be defined as one of the pre-calibrated C-arm poses on the source trajectory, e.g. the first one. Once $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ is known, the volumetric images are calibrated to the room via $^{\mathrm{W}}\mathbf{T}_{\mathrm{V}} = \,^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}(t_0)\,^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}\,^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}^{-1}(t_0)$, where t_0 denotes the time of calibration.

2.2   Prototype

For visualization of virtual content we use the Microsoft HoloLens (Microsoft, Redmond, WA) that simultaneously serves as inside-out tracker providing $^{\mathrm{W}}\mathbf{T}_{\mathrm{S}}$ according to see Sect. 2.1. To enable a master-slave configuration and enable tracking w.r.t. the exact same map of the environment, we mount a second HoloLens device on the C-arm to track movement of the gantry $^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}$. We use a CBCT enabled mobile C-arm (Siemens Arcadis Orbic 3D, Siemens Healthineers, Forchheim, Germany) and rigidly attach the tracking device to the image intensifier with the principal ray of the front facing RGB camera oriented parallel to the patient table as demonstrated in Fig. 2. $^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ is estimated via Hand-Eye calibration from 98 (tracker, C-arm) absolute pose pairs acquired during a circular source trajectory yielding 4753 relative poses. Since the C-arm is CBCT enabled, we simplify estimation of $^{\mathrm{V}}\mathbf{T}_{\mathrm{C}}$ and define t_0 to correspond to the first C-arm pose.

Fig. 2. The prototype uses a Microsoft Hololens as tracker which is rigidly mounted on the C-arm detector, as demonstrated in Fig. (a) and (b). In (c), the coordinate axis of the RGB tracker is shown in relation to the mobile C-arm.

Fig. 3. Since $^{\mathrm{S}}\mathbf{T}_{\mathrm{V}}(t)$ is known, the real object in (a) is overlaid with the rendered volume shown in (b). Correct overlay persists even if the real object is covered in (c).
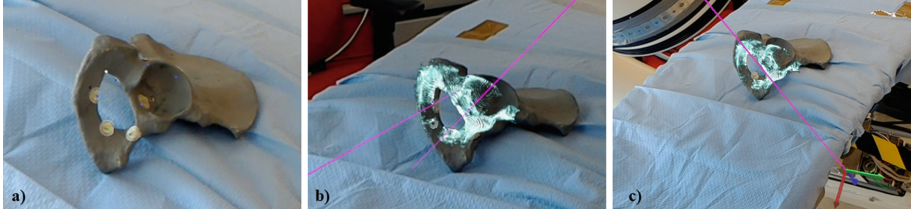

Fig. 4. (a) Pelvis phantom used for TRE assessment. (b) Lines placed during the experiment to evaluate point-to-line TRE. (c) Visualization of the X-ray source and principal ray next to the same phantom.

2.3   Virtual Content in the AR Environment

Once all spatial relations are estimated, multiple augmentations of the scene become possible. We support visualization of the following content depending on the task (see Fig. 4): Using $^{\mathrm{S}}\mathbf{T}_{\mathrm{V}}(t)$ we provide volume renderings of the 3D image volumes augmented on the patient's anatomy as shown in Fig. 3. In addition to the volume rendering, annotations of the 3D data (such as landmarks) can be displayed. Further and via $^{\mathrm{S}}\mathbf{T}_{\mathrm{C}}(t)$, the C-arm source and principal ray, seen in Fig. 4c) can be visualized as the C-arm gantry is moved to different viewing angles. Volume rendering and principal ray visualization combined are an effective solution to determine "bull's eye" views to guide punctures [11]. All rendering is performed on the HMD; therefore, the perceptual quality is limited by the computational resources of the device.

2.4   Experiments and Feasibility Study

Hand-Eye Residual Error: Following [8], we compute the rotational and translational component of $^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ independently. Therefore, we state the residual of solving $\mathbf{A}(t_i)\,^{\mathrm{T}}\mathbf{T}_{\mathrm{C}} = {}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}\mathbf{B}(t_i)$ for $^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$ separately for rotation and translation averaged over all relative poses.

Target Registration Error: We evaluate the end-to-end target registration error (TRE) of our prototype system using a Sawbones phantom (Sawbones, Vashon, WA) with metal spheres on the surface. The spheres are annotated in a CBCT

of the phantom and serve as the targets for TRE computation. Next, $M = 4$ medical experts are asked to locate the spheres in the AR environment: For every of the $N = 7$ spheres $\mathbf{p}_i$, the user j changes position in the room, and using the "air tap" gesture defines a 3D line $\mathbf{l}_i^j$ corresponding to his gaze that intersects the sphere on the phantom. The TRE is then defined as

$$\text{TRE} = \frac{1}{M \cdot N} \sum_{j=1}^{M} \sum_{i=1}^{N} \text{d}(\mathbf{p}_i, \mathbf{l}_i^j), \tag{3}$$

where $\text{d}(\mathbf{p}, \mathbf{l})$ is the 3D point-to-line distance.

Achieving "Bull's Eye" View: Complementary to the technical assessment, we conduct a clinical task-based evaluation of the prototype: Achieving "bull's eye" view for percutaneous punctures. To this end, we manufacture cubic foam phantoms and embed a radiopaque tubular structure (radius $\approx 5\,\text{mm}$) at arbitrary orientation but invisible from the outside. A CBCT is acquired and rendered in the AR environment overlaid with the physical phantom such that the tube is clearly discernible. Further, the principal ray of the C-arm system is visualized. Again, $M = 4$ medical experts are asked to move the gantry such that the principal ray pierces the tubular structure, thereby achieving the desired "bull's eye" view. Verification of the view is performed by acquiring an X-ray image. Additionally, users advance a K-wire through the tubular structure under "bull's eye" view guidance using X-rays from the view selected in the AR environment. Placement of the K-wire without breaching of the tube is verified in the guidance and a lateral X-ray view.

3   Results

Hand-Eye Residual Error: We quantified the residual error of our Hand-Eye calibration between the C-arm and tracker separately for rotational and translational component. For rotation, we found an average residual of $6.18°$, $5.82°$, and $5.17°$ around $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$, respectively, while for translation the root-mean-squared residual was $26.6\,\text{mm}$. It is worth mentioning that the median translational error in $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ direction was $4.10\,\text{mm}$, $3.02\,\text{mm}$, $43.18\,\text{mm}$, respectively, where $\mathbf{e}_z$ corresponds to the direction of the principal ray of the tracker coordinate system, i.e. the rotation axis of the C-arm.

Target Registration Error: The point-to-line TRE averaged over all points and users was $11.46\,\text{mm}$.

Achieving "Bull's Eye" View: Every user successfully achieved a "bull's eye" view in the first try that allowed them to place a K-wire without breach of the tubular structure. Fig. 5 shows representative scene captures acquired from the perspective of the user. A video documenting one trial from both a bystander's and the user's perspective can be found on our homepage.[1]
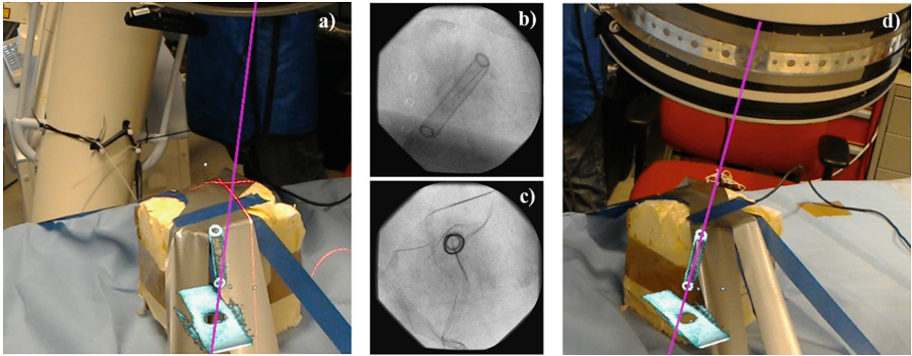
---

[1] https://camp.lcsr.jhu.edu/miccai-2018-demonstration-videos/.

Fig. 5. Screen captures from the user's perspective attempting to achieve the "bull's eye" view. The virtual line (purple) corresponds to the principal ray of the C-arm system in the current pose while the CBCT of the phantom is volume rendered in light blue. (a) The C-arm is positioned in neutral pose with the laser cross-hair indicating that the phantom is within the field of view. The AR environment indicates misalignment for "bull's eye" view that is confirmed using an X-ray (b). After alignment of the virtual principal ray with the virtual tubular structure inside the phantom (d), an acceptable "bull's eye" view is achieved (c).

4   Discussion and Conclusion

We presented an inside-out tracking paradigm to close the transformation loop for AR in orthopedic surgery based upon the realization that surgeon, patient, and C-arm can be calibrated to their environment. Our entirely marker-less approach enables rendering of virtual content at meaningful positions, i.e. dynamically overlaid with the patient and the C-arm source. The performance of our prototype system is promising and enables effective "bull's eye" viewpoint planning for punctures.

Despite an overall positive evaluation, some limitations remain. The TRE of 11.46 mm is acceptable for viewpoint planning, but may be unacceptably high if the aim of augmentation is direct feedback on tool trajectories as in [4,5]. The TRE is compound of multiple sources of error: (1) Residual errors in Hand-Eye calibration of ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{C}}$, particularly due to the fact that poses are acquired on a circular trajectory and are, thus, co-planar as supported by our quantitative results; and (2) Inaccurate estimates of ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{T}}$ and ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{S}}$ that indirectly affect all other transformations. We anticipate improvements in this regard when additional out-of-plane pose pairs are sampled for Hand-Eye calibration. Further, the accuracy of estimating ${}^{\mathrm{W}}\mathbf{T}_{\mathrm{T/S}}$ is currently limited by the capabilities of Microsoft's HoloLens and is expected to improve in the future.

In summary, we believe that our approach has great potential to benefit orthopedic trauma procedures particularly when pre-operative 3D imaging is available. In addition to the benefits for the surgeon discussed here, the proposed AR environment may prove beneficial in an educational context where

residents must comprehend the lead surgeon's actions. Further, we envision scenarios where the proposed solution can support the X-ray technician in achieving the desired views of the target anatomy.

References

1. Gay, B., Goitz, H.T., Kahler, A.: Percutaneous CT guidance: screw fixation of acetabular fractures preliminary results of a new technique with. Am. J. Roentgenol. **158**(4), 819–822 (1992)
2. Hong, G., Cong-Feng, L., Cheng-Fang, H., Chang-Qing, Z., Bing-Fang, Z.: Percutaneous screw fixation of acetabular fractures with 2D fluoroscopy-based computerized navigation. Arch. Orthop. Trauma Surg. **130**(9), 1177–1183 (2010)
3. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3D/2D registration methods for image-guided interventions. Med. Image Anal. **16**(3), 642–661 (2012)
4. Andress, S., et al.: On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. J. Med. Imaging **5** (2018)
5. Tucker, E., et al.: Towards clinical translation of augmented orthopedic surgery: from pre-op CT to intra-op X-ray via RGBD sensing. In: SPIE Medical Imaging (2018)
6. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W.: An evaluation of the RGB-D slam system. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 1691–1696. IEEE (2012)
7. Fotouhi, J., et al.: Pose-aware C-arm for automatic re-initialization of interventional 2D/3D image registration. Int. J. Comput. Assisted Radiol. Surg. **12**(7), 1221–1230 (2017)
8. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Trans. Rob. Autom. **5**(3), 345–358 (1989)
9. Berger, M., et al.: Marker-free motion correction in weight-bearing cone-beam CT of the knee joint. Med. Phys. **43**(3), 1235–1248 (2016)
10. De Silva, T., et al.: 3D–2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. Phys. Med. Biol. **61**(8), 3009 (2016)
11. Morimoto, M., et al.: C-arm cone beam CT for hepatic tumor ablation under real-time 3D imaging. Am. J. Roentgenol. **194**(5), W452–W454 (2010)