On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial

Sebastian Andress
Alex Johnson
Mathias Unberath
Alexander Felix Winkler
Kevin Yu
Javad Fotouhi
Simon Weidert
Greg Osgood
Nassir Navab

On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial

Sebastian Andress,[a,b,†] **Alex Johnson,**[c,†] **Mathias Unberath,**[a,*,†] **Alexander Felix Winkler,**[a,d,†] **Kevin Yu,**[a,d] **Javad Fotouhi,**[a] **Simon Weidert,**[b] **Greg Osgood,**[c] **and Nassir Navab**[a,d]
[a]Johns Hopkins University, Computer Aided Medical Procedures, Baltimore, Maryland, United States
[b]Ludwig-Maximilians-Universität München, Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie, Munich, Germany
[c]Johns Hopkins Hospital, Department of Orthopaedic Surgery, Baltimore, Maryland, United States
[d]Technische Universität München, Computer Aided Medical Procedures, Munich, Germany

Abstract. Fluoroscopic x-ray guidance is a cornerstone for percutaneous orthopedic surgical procedures. However, two-dimensional (2-D) observations of the three-dimensional (3-D) anatomy suffer from the effects of projective simplification. Consequently, many x-ray images from various orientations need to be acquired for the surgeon to accurately assess the spatial relations between the patient's anatomy and the surgical tools. We present an on-the-fly surgical support system that provides guidance using augmented reality and can be used in quasiunprepared operating rooms. The proposed system builds upon a multimodality marker and simultaneous localization and mapping technique to cocalibrate an optical see-through head mounted display to a C-arm fluoroscopy system. Then, annotations on the 2-D x-ray images can be rendered as virtual objects in 3-D providing surgical guidance. We quantitatively evaluate the components of the proposed system and, finally, design a feasibility study on a semianthropomorphic phantom. The accuracy of our system was comparable to the traditional image-guided technique while substantially reducing the number of acquired x-ray images as well as procedure time. Our promising results encourage further research on the interaction between virtual and real objects that we believe will directly benefit the proposed method. Further, we would like to explore the capabilities of our on-the-fly augmented reality support system in a larger study directed toward common orthopedic interventions. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.5.2.021209]

Keywords: interventional imaging; fluoroscopy; registration; surgical guidance.

Paper 17272SSRR received Sep. 15, 2017; accepted for publication Jan. 4, 2018; published online Jan. 26, 2018.

1 Introduction

Minimally invasive and percutaneous procedures with small incisions are an ongoing trend in orthopedic surgery.[1] As anatomic structures and location of tools and implants are not directly visible to the human eye, intraoperative imaging is needed for safe and effective procedures. The standard imaging modality still is fluoroscopic imaging by a C-arm device, displaying two-dimensional (2-D) x-ray projection images to the surgeon on a separate monitor. The most demanding tasks using fluoroscopy are those involving the precise placement of tools or implants. Due to the 2-D projection, the position of an instrument's tip, for instance on the axis perpendicular to the image plane, cannot be safely determined. To compensate for this, additional images have to be acquired from another angle, ideally 90 deg to create two-planar images that allow the surgeon to estimate the true three-dimensional (3-D) position of any radiopaque structure within the field of view.[2] The mental projection of 2-D images onto a 3-D world often is counterintuitive and error-prone due to projective simplification and high mental work-load. The result often is repetitive x-ray imaging to control the procedure, increasing the radiation dose for the patient but especially for the surgeon who is often standing very close to the patient with his hands close to the beam.[3,4] Another problem is that fluoroscopic imaging is noncontinuous but rather provides single snapshots, leaving the surgeon "blind" in-between image acquisitions. Both, projection errors and noncontinuous imaging can lead to misplacement of tools or implants, potentially leading to tissue damage with consequences ranging from hematoma and bleeding to nerve damage or joint destruction.[5,6] The potential results of this are poor outcomes and revision rates.

Computer aided surgery (CAS), also known as computer navigation has formerly been introduced to offer the surgeon a more powerful visualization of his actions within the 3-D space, continuously displaying tracked instruments or implants within the patients anatomy mostly using preexisting CT or MRI datasets. Those systems are used in a wide range of procedures ranging from neurosurgery to orthopedic surgery. All of those systems require an initial registration process and calibration of specialized tools. In most cases, the patient as well as the instruments is tracked by optical markers attached to them visible to infrared cameras. An abundance of the literature exists on the benefits of these systems on accuracy and safety.[7] But other tracking solutions such as electromagnetic tracking have also shown the ability to significantly reduce radiation without substantially increasing surgical times[8] while introducing further requirements as an electromagnetically shielded operating field. Despite the positive effects described, CAS still is far from dominating the operating room. The reasons for this being not only the high cost of investment but also the added

---

*Address all correspondence to: Mathias Unberath, E-mail: unberath@jhu.edu

†These authors have contributed equally and are listed in alphabetical order.

surgery time caused by technical setup and additional workflow steps, such as repeated registration. Studies saw an increase in total procedure time up to 65 min without clearly improving clinical outcomes.[9–13] Furthermore, those systems are often associated with a long learning curve until an efficient workflow is established and the team learned how to prevent or deal with technical difficulties such as registration errors and line-of-sight problems while tracking.[14] Another reason clearly is the distraction caused by displaying visualization on a separate monitor far from the real action and in a totally different coordinate system as the real world, limiting ergonomics and usability especially for the new user.[15] Therefore, there is much potential for improvement and new technologies that offer a more intuitive approach and facilitate percutaneous surgery without introducing a complicated setup or requiring major changes to the operative workflow.

Augmented reality (AR) is a technology that promises to integrate computer guidance into the surgical workflow in a very intuitive way by providing visual guidance directly related to the anatomical target area inside the patient in front of the surgeons eyes. Current CAS systems usually display information on 2-D monitors mounted on carts or the ceiling of the theater.[16,17] As the relation between the monitor and the patient or the surgeon is unknown to the system, it displays a view onto the anatomy as well as tool trajectories that are unrelated to the real perspective of the surgeon. AR systems to support orthopedic surgery have been demonstrated and evaluated already. One of the most studied designs consists of a C-arm with a calibrated video camera attached to it, thus being able to augment the live video image with fluoroscopic images in precise overlay.[16] Its most recent variation relies on an RGBD camera rigidly mounted on the C-arm detector and calibrated to an intraoperative cone-beam CT (CBCT) scan. It has the ability to simultaneously render multiple digitally reconstructed radiographs at different viewing angles overlaid with a 3-D cloud of points from the RGBD camera showing the surgeon's hands and tools.[18,19] Despite promising performance that includes remarkable reductions in both surgery time and dose without substantial changes to the traditional workflow, the system requires an intraoperative CBCT scan, which is only available on high-end C-arm systems and adds operation time and dose. Replacing the CBCT with 2-D/3-D registration of preoperative CT to interventional fluoroscopy mitigates aforementioned challenges and makes the system more usable for many orthopedic surgical applications.[20,21] The use of HMD for AR in orthopedic surgery is suggested in the literature for guiding the placement of percutaneous sacroiliac screws in pelvic fractures.[22] This approach relies on external navigation systems to track the drill, pelvis, and HMD.

Within this paper, we propose an easy-to-use guidance system with the specific aim of eliminating potential roadblocks to its use regarding system setup, change in workflow, or cost. The system is applicable to most fluoroscopy-guided orthopedic surgeries without the need for 3-D pre- or intraoperative imaging and provides support for surgeon's actions through an AR environment based on optical see-through HMD that is calibrated to the C-arm system. The proposed solution eliminates the need for external navigation hardware, as well as the preoperative calibration of the sensors. It allows visualizing the path to anatomical landmarks annotated in x-ray images in 3-D directly on the patient. Calibration of intraoperative fluoroscopy imaging to the AR environment is achieved on-the-fly using a mixed-modality

fiducial that is imaged simultaneously by the HMD and the C-arm system. Therefore, the proposed system effectively avoids the use of dedicated but impractical optical or electromagnetic tracking solutions with 2-D/3-D registration, complicated setup, or use of which is associated with the most substantial disruptions to the surgical workflow. The aim of this study is to describe the calibration of the system and to determine its performance in a set of experiments carried out by expert surgeons. The first experiment will determine accuracy and precision when using a technical phantom consisting of spherical radiopaque target points that are surrounded by soft-tissue. The second experiment mimics a typical step in many surgical procedures, percutaneously placing a K-wire onto a specific location of the patients' anatomy. In our case, the task will be finding the entry-point for an intramedullary nail at the tip of the greater trochanter. This task is typical for most percutaneous orthopedic procedures such as pedicle screw, lag screw placement, or interlocking of nails.

The paper is outlined as follows. In Sec. 1, we discussed problems of the classic fluoroscopic guidance technique during minimal invasive surgeries. A concrete example describes the need for another guidance technique. Current solutions in different research fields are addressed, leading to the gap of a simple on-the-fly guidance technique for surgeries. Section 2 provides an overview of the presented solution. In the following, the used devices and their spatial transformations are explained. A multimodality marker is introduced for the registration of the system using visual marker tracking. We then discuss the C-arm and its integration with the system. Next, we describe how HMD devices are used together with the C-arm scanner to enable real-time AR. Four experiments are then presented to evaluate the accuracy and precision of each individual component, followed by a summary of the results in Sec. 3. Lastly, we discuss the experimental results and provide a summary of the proposed AR solutions in Secs. 4 and 5, respectively.

2 Method

2.1 *System Overview*

The proposed system comprises three components that must exhibit certain characteristics to enable on-the-fly AR guidance: a mixed-modality fiducial, a C-arm x-ray imaging system, and an optical see-through HMD. We will use the following notation of expressing transformations: The transformation $^A\mathbf{T}_B$ is defined as the transformation from coordinate system A to coordinate system B. This notation enables to concatenate transformations easily as in $^A\mathbf{T}_C = {}^B\mathbf{T}_C{}^A\mathbf{T}_B$. Based on these components, the spatial relations that need to be estimated in order to enable real-time AR guidance are shown in Fig. 1. Put concisely, we are interested in recovering the transformation $^C\mathbf{T}_{HMD}(t)$ that propagates information from the C-arm to HMD coordinate system while the surgeon moves over time t. To this end, we need to estimate the following transformations:

- $^C\mathbf{T}_M$: Extrinsic calibration of the C-arm to the multimodality marker domain, obtained from the x-ray image using ARToolKit (see Sec. 2.2).

- $^{HMD}\mathbf{T}_M$: Transformation describing the relation between the HMD and the multimodality marker coordinate system, estimated from the RGB image acquired by the HMD using ARToolKit (Sec. 2.2).
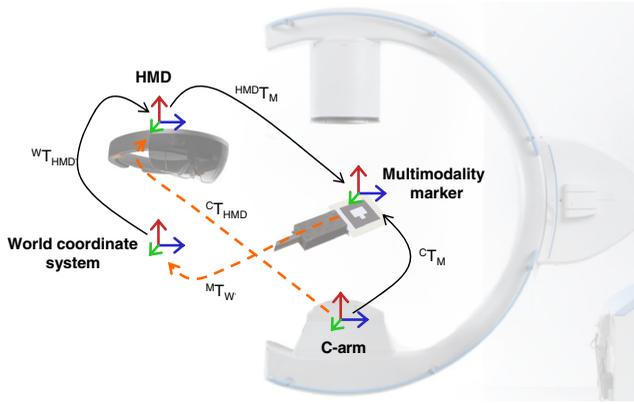
Fig. 1 Spatial transformations for the on-the-fly AR solution.

- $^W\mathbf{T}_{HMD}$: The HMD is capable of establishing a map of its surroundings with arbitrary origin while localizing itself therein. $^W\mathbf{T}_{HMD}$ then describes the pose of the HMD within this so-called world coordinate system. In practice, it is computed using vision-based tracking algorithms such as simultaneous localization and mapping (SLAM)[23] (Sec. 2.4).

- $^M\mathbf{T}_W$: Describes the mapping from the multimodality marker to the world coordinate system. It is estimated using $^W\mathbf{T}_{HMD}$ and $^{HMD}\mathbf{T}_M$ in the calibration phase (see Sec. 2.4).

Once these relations are known, annotations in an intraoperatively acquired x-ray image can be propagated to and visualized by the HMD, which provides support for placement of wires and screws in orthopedic interventions. The transformation needed is given as

$$^C\mathbf{T}_{HMD}(t) = {}^W\mathbf{T}_{HMD}(t)\underbrace{[^W\mathbf{T}_{HMD}^{-1}(t_0){}^{HMD}\mathbf{T}_M^{-1}(t_0)]^C\mathbf{T}_M(t_0)}_{^C\mathbf{T}_W},$$

$$(1)$$

where t_0 denotes the time of calibration, e.g., directly after repositioning of the C-arm, suggesting that $^C\mathbf{T}_W$ is constant as long as the C-arm remains in place. For brevity of notation, we will omit the time dependence of the transformations whenever they are clear or unimportant.

We provide detailed information on the system components and how they are used to estimate aforementioned transformations in the following sections.

2.2 *Multimodality Marker*

The key component of the proposed system is a multimodality marker that can be detected using C-arm as well as the HMD using x-ray and RGB imaging devices, respectively. As the shape and size of the multimodality marker is precisely known in 3-D, estimation of both transforms $^C\mathbf{T}_M$ and $^{HMD}\mathbf{T}_M$ is possible in a straightforward manner if the marker can be detected in the 2-D images. To this end, we rely on the well-known ARToolKit for marker detection and calibration[24] and design our multimodality marker accordingly.

The marker needs to be well discernible when imaged using the optical and x-ray spectrum. To this end, we 3-D print the

template of a conventional ARToolKit marker, as shown in Fig. 2(a), that serves as the housing for the multimodality marker. Then, we machined a metal inlay (solder wire 60\40 Sn\Pb) that strongly attenuates x-ray radiation, see Fig. 2(b). After covering the metal with a paper printout of the same ARToolKit marker as shown in Fig. 2(c), the marker is equally well visible in the x-ray spectrum as well as RGB images due to the high attenuation of lead as can be seen in Fig. 2(d). This is very convenient, as the same detection and calibration pipeline readily provided by ARToolKit can be used for both images.

It is worth mentioning that the underlying vision-based tracking method in ARToolKit is designed for reflection and not for transmission imaging, which can be problematic in two ways. First, ARToolKit assumes 2-D markers suggesting that the metal inlay must be sufficiently thin in order not to violate this assumption. Second, a printed marker imaged with an RGB camera perfectly occludes the scene behind it and is, thus, very well visible. For transmission imaging, however, this is not necessarily the case as all structures along a given ray contribute to the intensity at the corresponding detector pixel. If other strong edges are present close to this hybrid marker, detection and hence calibration may fail. To address both problems simultaneously we use digital subtraction, a concept that is well known from angiography.[25,26] We acquire two x-ray images using the same acquisition parameters and C-arm pose both with and without the multimodality marker introduced into the x-ray beam. Logarithmic subtraction then yields an image that, ideally, only shows the multimodality marker and lends itself well to marker detection and calibration using the ARToolKit pipeline. Moreover, subtraction imaging allows for the use of very thin metal inlays as subtraction artificially increases the contrast achieved by attenuation only. While the subtraction image is used for processing, the surgeon is shown the fluoroscopy image without any multimodality marker obstructing the scene.

2.3 *C-Arm Fluoroscopy System*

The proposed system has the substantial advantage that, in contrast to many previous systems,[19,27] it does not require any modifications to commercially available C-arm fluoroscopy systems. The only requirement is that images acquired during the intervention can be accessed directly such that geometric calibration is possible. Within this work, we use a Siemens ARCADIS Orbic 3D (Siemens Healthcare GmbH, Forchheim, Germany) to acquire fluoroscopy images and a frame grabber (Epiphan Systems Inc., Palo Alto, California) paired with a streaming server[15] to send them via a wireless local network to the HMD.

While extrinsic calibration of the C-arm system is possible using the multimodality marker as detailed in Sec. 2.2, the intrinsic parameters of the C-arm, potentially at multiple poses, are estimated in a one-time offline calibration, e.g., as described by Fotouhi et al.[28] using a radiopaque checkerboard.

Once the extrinsic and intrinsic parameters are determined, the 3-D source and detector pixel positions can be computed in the coordinate system of the multimodality marker. This is beneficial, as simple point annotations on the fluoroscopy image now map to lines in 3-D space that represent the x-ray beam emerging from the source to the respective detector pixel. These objects, however, cannot yet be visualized at a meaningful position as the spatial relation of the C-arm to the HMD is unknown. The multimodality marker enabling calibration must be imaged simultaneously by the C-arm system and the
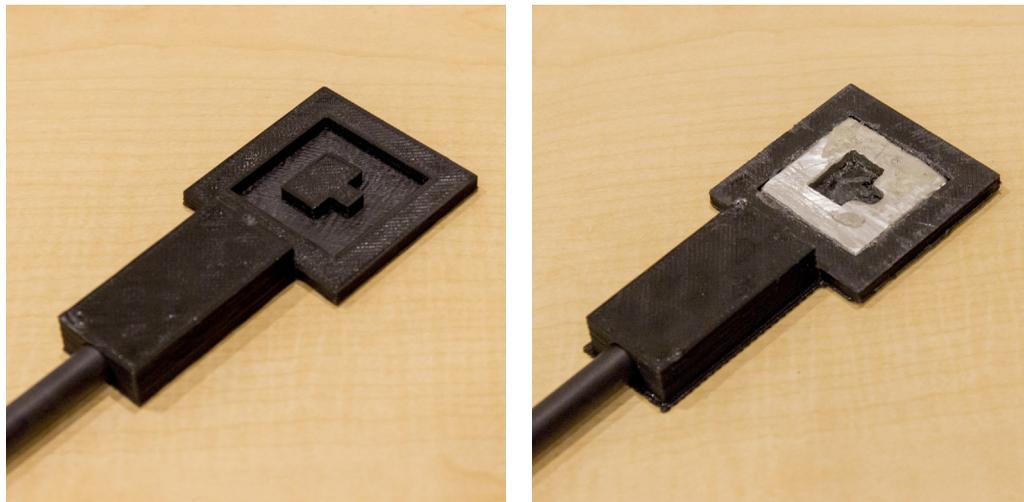
(a) Template of the multimodality marker after 3-D printing.

(b) 3-D printed template filled with metal to create a radiopaque multimodality marker.
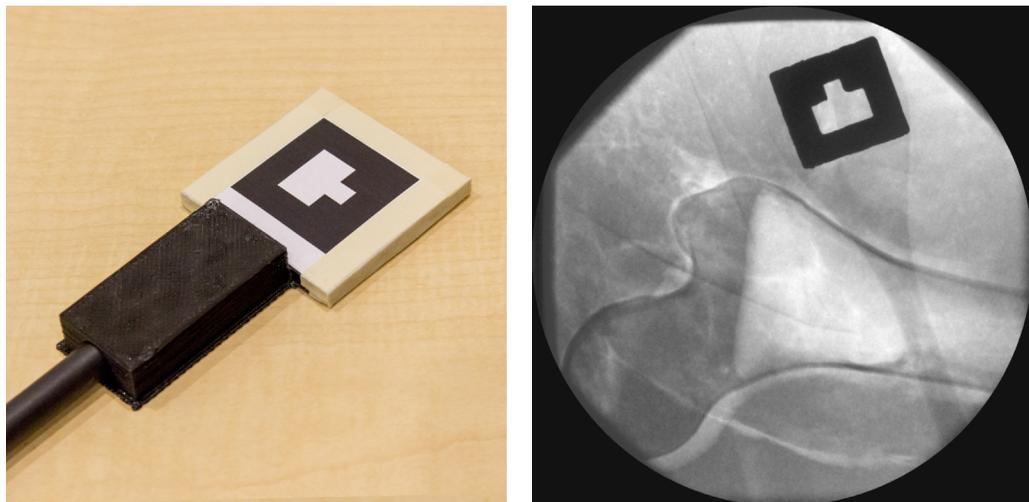
(c) Radiopaque marker overlaid with printout of the same ARToolKit marker to increase contrast for optical imaging.

(d) X-ray intensity image of the proposed multimodality marker. Due to the high attenuation of lead, the ARToolKit marker appears similar when imaged in the X-ray or optical spectrum.

Fig. 2 Steps in the creation of the multimodality marker. The 3-D-printed template serves as a housing for the marker and is rigidly attached to a carbon fiber rod such that the marker can be safely introduced into the x-ray field of view.

RGB camera on the HMD to enable meaningful visualization in an AR environment. This process will be discussed in greater detail below.

2.4 Optical See-Through HMD and the World Coordinate System

The optical see-through HMD is an essential component of the proposed system as it needs to recover its pose with respect to the world coordinate system at all times, acquire and process optical images of the multimodality marker, allow for interaction of the surgeon with the supplied x-ray image, combine and process the information provided by the surgeon and the C-arm, and provide real-time AR visualization for guidance. Within this work, we rely on the Microsoft HoloLens (Microsoft Corporation, Redmond, Washington) as the optical

see-through HMD as its performance compared favorably to other commercially available devices.[29]

2.4.1 Pose estimation

Similar to the pose estimation for the C-arm, we first seek to estimate the pose of the HMD with respect to the multimodality marker $^{HMD}\mathbf{T}_M$. In order to allow for calibration of the C-arm to the HMD, the images of the marker used to retrieve $^{C}\mathbf{T}_M$ and $^{HMD}\mathbf{T}_M$ for the C-arm and the HMD, respectively, must be acquired with the marker at the same position. If the multimodality marker is handheld, the images should ideally be acquired at the same time t_0. The HoloLens is equipped with an RGB camera that we use to acquire an optical image of the multimodality marker and estimate $^{HMD}\mathbf{T}_M$ using ARToolKit as described in Sec. 2.2.

In principle, these two transformations are sufficient for AR visualization but the system would not be appropriate: if the surgeon wearing the HMD moves, the spatial relation $^{HMD}\mathbf{T}_M$ changes. As limiting the surgeon's movements is not feasible, updating $^{HMD}\mathbf{T}_M(t)$ over time may seem like an alternative but is impracticable as it would require the multimodality marker to remain at the same position, potentially close to the operating field. While updating $^{HMD}\mathbf{T}_M(t)$ over time seems complicated, recovering $^{HMD}\mathbf{T}_W(t)$, the pose of the HMD with respect to the world coordinate system, is readily available from the HoloLens HMD and is estimated using a proprietary algorithm based on concepts similar to SLAM.[23,30,31] Consequently, rather than directly calibrating the C-arm to the HMD, we calibrate the C-arm to the world coordinate system (in the HoloLens community sometimes referred to as world anchor or spatial map) to retrieve $^C\mathbf{T}_W$ that is constant if the C-arm is not repositioned.

2.4.2 User interface and AR visualization

In order to use the system for guidance, key points must be identified in the x-ray images. Intraoperative fluoroscopy images are streamed from the C-arm to the HMD and visualized using a virtual monitor as described in greater detail in Ref. 15. The surgeon can annotate anatomical landmarks in the x-ray image by hovering the HoloLens cursor over the structure and performing the air tap gesture. In 3-D space, these points must lie on the line connecting the C-arm source position and the detector point that can be visualized to guide the surgeon using the spatial relation in Eq. (1). An exemplary scene of the proposed AR environment is provided in Fig. 3. Guidance rays are visualized as semitransparent lines with a thickness of 1 mm while the C-arm source position is displayed as a cylinder. The association from annotated landmarks in the x-ray image to 3-D virtual lines is achieved via color coding.

It is worth mentioning that the proposed system allows for the use of two or more C-arm poses simultaneously. When two views are used, the same anatomical landmark can be annotated in both fluoroscopy images allowing for stereo reconstruction of the landmark's 3-D position.[32] In this case, a virtual sphere is shown in the AR environment at the position of the triangulated 3-D point, shown in Fig. 3(b). Furthermore, the interaction allows for the selection of two points in the same x-ray image that define a line. This line is then visualized as a plane in the AR environment. An additional line in a second x-ray image can be annotated resulting in a second plane. The intersection of these two planes in the AR space can be visualized by the surgeon and followed as a trajectory.

2.5 Integration with the Surgical Workflow

As motivated in Sec. 1, one of the main goals of this study was to create an easy on-the-fly guidance system. The simple setup proposed here is enabled by the multimodality marker and, more substantially, by the capabilities of the HoloLens.

Configuring the system in a new operating room requires access to the C-arm fluoroscopy images and setup of a local wireless data transfer network. Once the HMD is connected

(a) AR environment with a single C-arm view.

(b) AR environment when two C-arm views are used.
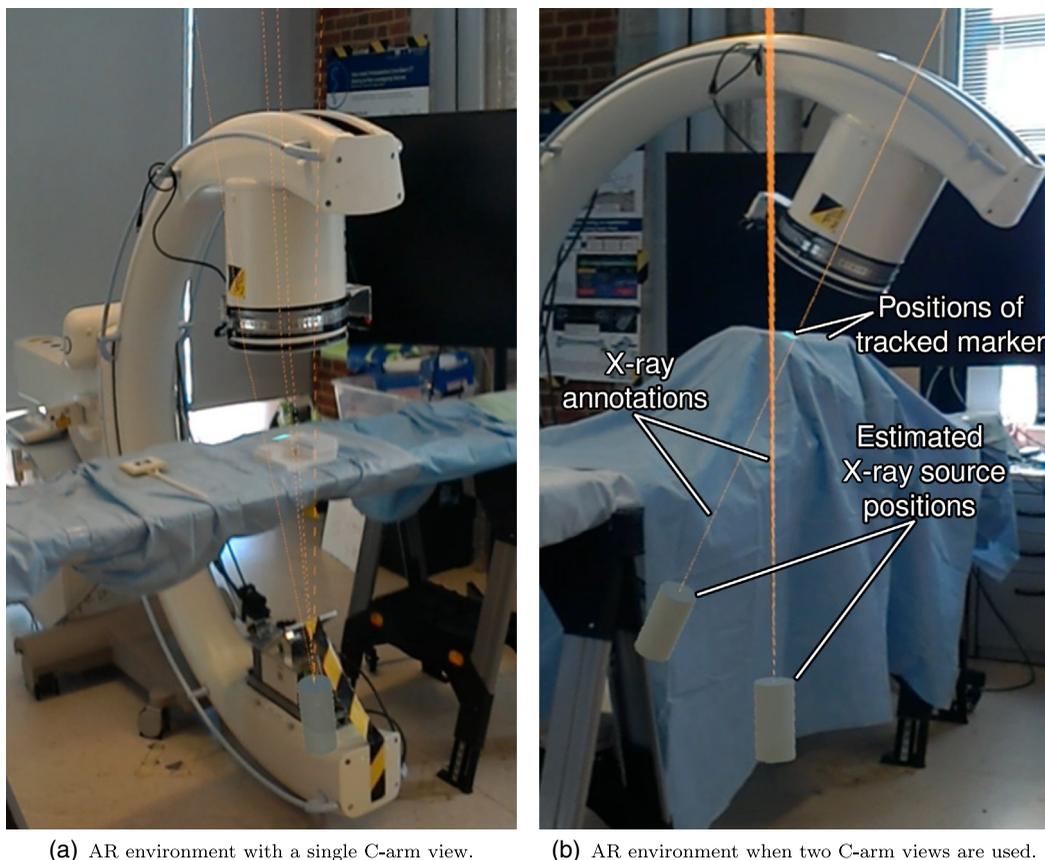
Fig. 3 Source position of the C-arm shown as a cylinder and virtual lines that arise from annotations in the fluoroscopy image.

to the C-arm, only very few steps for obtaining AR guidance are needed. For each C-arm pose, the surgeon has to

1. Position the C-arm using the integrated laser cross-hair such that the target anatomy will be visible in fluoroscopy.

2. Introduce the multimodality marker in the C-arm field of view and also visible in the RGB camera of the HMD. If the fiducial is recognized by the HMD, an overlay will be shown. Turning the head such that the marker is visible to the eye in straight gaze is usually sufficient to achieve marker detection.

3. Calibrate the system by use of a voice command (Lock) and simultaneously acquiring an x-ray image with the marker visible in both modalities. This procedure defines t_0 and thus $^C\mathbf{T}_W$ in Eq. (1). Note that in the current system, a second x-ray images need to be acquired for subtraction (see Sec. 2.2) but the marker can now be removed from the scene.

4. Annotate the anatomical landmarks to be targeted in the fluoroscopy image as described in Sec. 2.4.

Performing the aforementioned steps yields virtual 3-D lines that may provide sufficient guidance in some cases; however, the exact position of the landmark on this line remains ambiguous. If the true 3-D position of the landmark is needed, the above steps can be repeated for another C-arm pose.

2.6 Experiments

We design experiments to separately evaluate the system's components quantitatively. While the first two studies do not require user interaction and objectively assess system performance, the last two experiments are designed as a preliminary feasibility study that is performed by two orthopedic surgeons at the Johns Hopkins Hospital.

2.6.1 Calibration

In the first experiment, we seek to assess how well the multimodality marker enables calibration of the system. To this end, the HMD and C-arm remain at fixed poses and are calibrated using the procedure described in Sec. 2.4 yielding

$$^C\mathbf{T}_{HMD}(t_0) = {}^{HMD}\mathbf{T}_M^{-1}(t_0){}^C\mathbf{T}_M(t_0). \tag{2}$$

This does not involve the world coordinate system as the pose of HMD with respect to the C-arm is constant over time. Then, the multimodality marker is displaced multiple times and each time $^C\mathbf{T}_{HMD}(t_i)$ is updated, where $i = 1, \ldots, 6$. As the spatial relation between the C-arm and the HMD remains unchanged, disagreement of $^C\mathbf{T}_{HMD}(t_i)$ and $^C\mathbf{T}_{HMD}(t_0)$ is related to calibration performance and reproducibility. We repeat this procedure six times and report the mean and standard deviation of positional and rotational errors.

2.6.2 HMD tracking

The transformation $^C\mathbf{T}_{HMD}(t)$ is time dependent as the surgeon is free to move and, thus, $^{HMD}\mathbf{T}_W(t)$ changes. Accurate estimation of $^{HMD}\mathbf{T}_W(t)$ is crucial to ensure that the virtual objects designated for surgical guidance are displayed at the correct

position on the patient. To evaluate the tracking performance of the HMD, i.e., the HoloLens, the multimodality marker is fixed to the surgical bed while the HMD is mounted facing the multimodality marker on a tripod to avoid inaccuracies due to shaking or very fast movements. We obtain reference 3-D positions of the corner points of the multimodality marker in the world coordinate system via

$$^M\mathbf{T}_W(t_0) = {}^W\mathbf{T}_{HMD}^{-1}(t_0){}^{HMD}\mathbf{T}_M^{-1}(t_0) \tag{3}$$

and then reposition the HMD $i = 1, \ldots, 6$ times yielding estimates of $^M\mathbf{T}_W(t_i)$ and, thus, the 3-D corner points of the multimodality marker. We repeat this procedure six times and report the root-mean-square error over all corner points.

2.6.3 Precision of 3-D landmark identification

We assess the precision of 3-D landmark retrieval when the system is used in two-view mode. To this end, we construct a staircase phantom with a metal bead attached to each plateau [see Fig. 4(a)]. In this scenario, the metal beads serve as landmarks for annotation in an otherwise featureless image. We image the phantom in two C-arm gantry positions according to the workflow outlined in Sec. 2.5. The first C-arm position corresponds to a view where the detector is parallel to the baseplate of the phantom, while the second view is rotated by ∼15 deg. The 3-D positions of the four landmark points in the C-arm coordinate frame are computed via triangulation from the respective corresponding annotations in the two x-ray images. The experiment is repeated five times and the system is recalibrated using the multimodality marker. To assess the precision, and thus reproducibility, of the calibration, we compute the centroid and standard deviation for each of the four metal spheres. Here, centroid refers to the mean position among all corresponding landmarks. We then state the average standard deviation in millimeter as a measure for precision.

2.6.4 Guidance using the AR environment

The proposed system provides guidance to the surgeon by rendering a virtual line that passes through the desired target that is annotated in an x-ray image. Assuming ideal calibration of the C-arm as well as perfect localization of the HMD in the world coordinate system, it is unclear how well tools, such as K-wires, can be aligned with the virtual objects. This problem is emphasized as important cues, such as occlusion, are not yet modeled for the interaction of real with virtual objects.

To establish an upper bound on the guidance performance, we use the staircase phantom introduced above. In the user study, the phantom is positioned in the center of the C-arm field of view such that the base is parallel to the detector plane and then covered using a layer of ballistic gel. Using the multimodality marker, the C-arm is calibrated to the HMD and all four metal beads are annotated in an x-ray image by the user yielding four virtual guidance lines. The user is then asked to follow each line with a K-wire penetrating the ballistic gel and piercing the phantom. This targeting procedure is performed in a clockwise manner 20 times (five times per marker). We then compute the average distance of the punctures to the true metal bead locations.

2.6.5 *Semianthropomorphic femur phantom*

Finally, we assess the complete system performance in a user study on a semianthropomorphic femur phantom simulating entry point localization for the implantation of cephalomedullary nails to treat proximal femoral fractures. The phantom is shown in Fig. 4(b). In order to reduce ambiguity of the desired entry point, the target is defined to be the tip of the greater trochanter that can be well perceived and annotated in the x-ray images. The phantom simulates an obese patient such that the femur is encased in a thick envelope of ballistic gel. The gel casing and the femur inside are positioned on the table in a manner consistent with a real surgical scenario. The users are asked to navigate a K-wire onto the target point first, using the conventional approach without guidance and second, using the proposed system in the two-view scenario: as opposed to the previous experiment the participants do not rely on only a single projective x-ray image that they can annotate but also are allowed to annotate two x-ray images. Annotating the same anatomical point in both images results in two guidance lines intersecting at the point of interest. This point can then serve as the target from an arbitrary incision point. After the K-wire is placed, its distance to the desired position is assessed in 3-D using CBCT. Moreover, we record and report the procedure time and the amount of x-ray images acquired.

3 Results

3.1 *Calibration*

To measure the precision and robustness of the system calibration, we estimate the poses of the multimodality marker from a static HMD at different locations. The overall average of Euclidean distances to the centroid of measurements (positional error) between $^C\mathbf{T}_{\mathrm{HMD}}(t_i)$ and $^C\mathbf{T}_{\mathrm{HMD}}(t_0)$ is 21.4 mm with a standard deviation of 11.4 mm. The overall average of angles to the average orientation (rotational error) is 0.9 deg with a standard deviation of 0.4 deg. It is important to note that, as
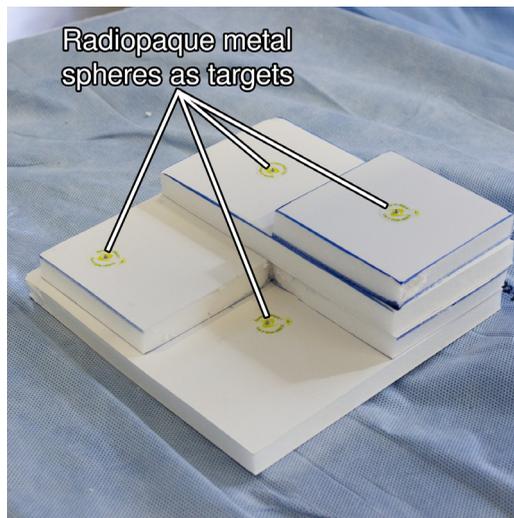
(a) Staircase phantom.

(b) Semianthropomorphic femur phantom. An artificial bone model is encased in gel which is then covered with a plastic bag.
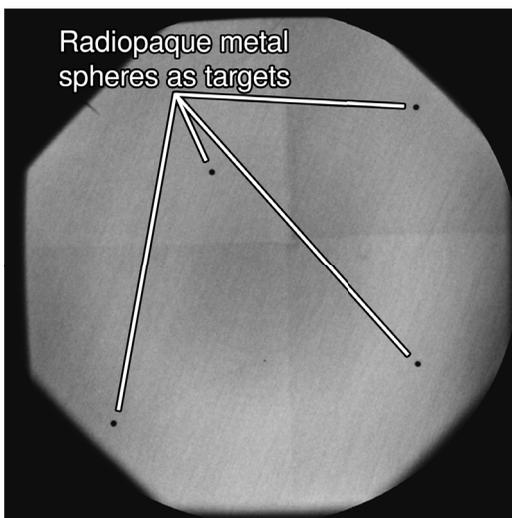
(c) Staircase phantom in X-ray.

(d) Semianthropomorphic femur phantom in X-ray.

Fig. 4 Phantoms used in the user studies assessing the performance of the system in an isolated and a surgery-like scenario in (a) and (b), and in x-ray (c) and (d), respectively.

Table 1 Deviations of estimated 3-D landmark positions from the respective centroid. The average distance is stated as a tuple of mean and standard deviation. All values are stated in millimeters.

	First run	Second run	Third run	Fourth run	Fifth run	Average
Target P1	19.3	3.54	4.78	8.31	11.6	(9.49,6.31)
Target P2	12.9	4.22	7.67	11.5	7.56	(8.76,3.44)
Target P3	18.2	4.19	6.76	10.4	6.29	(9.18,5.53)
Target P4	21.6	6.62	3.51	18.8	8.23	(11.7,7.93)

there is no ground-truth available for C-arm x-ray poses in this experiment, we only report the consensus between measurements, i.e., the precision of the calibration step.

3.2 HMD Tracking

In the experiment evaluating the HMD tracking accuracy, we found a root-mean-square error of 16.2 mm with a standard deviation of 9.5 mm. As described in Sec. 2.6, the error is measured among all the corner points of the multimodality marker.

3.3 Precision of 3-D Landmark Identification

As stated in Table 1, the average distance to the centroids ranged from 8.76 to 11.7 mm. The in-plane error, evaluated by projecting the 3-D deviations onto the detector plane of the first x-ray orientation is substantially lower and ranges from 3.21 to 4.03 mm, as summarized in Table 2.

3.4 Guidance Using the AR Environment

The experiment performed by two expert users on step phantoms as shown in Fig. 4(a) resulted in an average precision error of 4.47 mm with a standard deviation of 2.91 mm measured as the average Euclidean distance to the centroid of the puncture marks. The accuracy of this system is then measured as the average distance to metal beads, i.e., ground-truth, which yielded 9.84-mm error with a standard deviation of 3.97 mm. The mean errors and standard deviations of both participants are summarized in Table 3, showing the average distance of each attempt to the centroid of all attempts and the average distance of this centroid to the true metal bead target. Figure 5(b) shows the target and the centroid on the step phantom.

3.5 Semianthropomorphic Femur Phantom

For this experiment, surgeons were asked to perform a simulated K-wire placement on the semianthropomorphic femur phantom with the on-the-fly AR system, as well as classic fluoro-guided approach. The average distance of the tip of the K-wire to the tip of the greater trochanter is 5.20 mm with the proposed AR solution and 4.60 mm when only fluoroscopic images were used. However, when the proposed solution was used, the average number of x-ray images substantially decreased. The participants needed five x-ray acquisitions from two orientations and on average 16 x-ray images from six orientations when the proposed and traditional solution were used, respectively. In fact, the number of images for our solution can further be decreased by two as we include the images required for background subtraction that may become obsolete with a different marker design. Finally, the procedure time reduced from 186 s (standard deviation of 5 s) in the classic approach to 168 s (standard deviation of 18 s). An x-ray of the K-wire at the final position of one of the participants can be seen in Fig. 6.

Table 2 Deviations of estimated 3-D landmark positions from the respective centroid projected onto the x-ray plane of the first view. Again, the average distance is stated as a tuple of mean and standard deviation and all values are given in millimeters.

Distance to	First run	Second run	Third run	Fourth run	Fifth run	Average
Centroid P1	4.59	3.42	4.78	0.32	2.96	(3.21,1.79)
Centroid P2	5.05	3.71	6.06	3.17	0.36	(3.67,2.17)
Centroid P3	3.93	3.82	5.11	3.32	3.57	(3.96,0.70)
Centroid P4	6.48	4.24	3.45	4.69	1.31	(4.03,1.89)

Table 3 Errors measured as distances to the target and the centroid. All values are stated in millimeters in tuples of mean and standard deviation. Results are shown for both participants (P1 and P2).

Distance to	First target	Second target	Third target	Fourth target	Overall
Target P1	(8.17,3.85)	(8.08,1.63)	(7.97,4.08)	(5.65,1.56)	(7.47,3.20)
Centroid P1	(5.93,4.02)	(2.49,1.29)	(5.84,2.40)	(2.49,1.22)	(4.19,3.03)
Target P2	(12.3,2.1)	(11.3,0.8)	(15.2,3.9)	(10.1,2.4)	(12.2,3.2)
Centroid P2	(4.92,2.96)	(4.54,2.80)	(6.30,2.80)	(3.25,1.27)	(4.75,2.77)

(a) AR view of the step phantom augmented with annotations. Note that the world anchored marker has shifted from its tracked position after locking it in place. This results in a shift in annotations shown in orange color.

(b) Photo of one target used to evaluate the guidance accuracy: Note that the puncture marks are closely clustered, the centroid is however not aligned with the target metal bead.

Fig. 5 Setup and the AR view for the "Guidance Experiment Using the AR Environment."

4 Discussion

4.1 Outcome of the Preliminary Feasibility Studies

The experiments conducted in this paper are designed to distinguish between accuracy and precision errors of three different parts of the proposed AR support system: the calibration between the RGB camera of the HMD and the C-arm, the world tracking of the HMD, and the visualization of the guidance.

The results of the calibration experiment, where the C-arm source is tracked with respect to the HMD, indicate large positional error and low orientational error. Two main sources for this error are (i) error propagation due to large distances among the HMD, marker, and the x-ray source where small errors in marker tracking translate to large displacements in the estimation of the pose of the x-ray source and (ii) errors in marker tracking that increase when the multimodality marker is not facing parallel to the RGB camera on the HMD. The HMD Tracking experiment indicates a drift in tracking of the multimodality marker with respect to the world anchor as the user observes the marker from different locations during the intervention. This error decreases in a static environment where the spatial map of the HMD works more reliably. All aforementioned sources of error affect the reproducibility reported in the 3-D landmark identification experiment. The distance from the centroid was substantially reduced when only the in-plane error was considered, an observation that is well explained by the narrow baseline between the two x-ray poses. Yet, the in-plane distance found in this experiment is in good agreement with the precision reported for the user study on a similar step-phantom.

The quantitative error measures reported in Sec. 3 suggest lackluster performance of some of the subsystem components that would inhibit clinical deployment for procedures where very high accuracy is paramount. However, in scenarios where rough guidance is acceptable, the overall system performance evaluated on the semianthropomorphic femur phantom is promising. The distance of the K-wire from the anatomical landmark is comparable, yet, the proposed system requires the acquisition of fewer x-ray images. The results suggest that the proposed on-the-fly AR solution may already be adequate to support surgeons in bringing surgical instruments close to the desired anatomical landmarks. It is worth mentioning that many of the limitations discussed here are imposed on our prototype solution as it relies on currently available hardware or software. Consequently, improvements in these devices will directly benefit the proposed workflow for on-the-fly AR in surgery.

4.2 Challenges

Our method combines two approaches to create an easy guidance system. First, it utilizes the accessible tracking capability of the HMD, to use the spatial map and its world anchor as the fundamental coordinate system of the tracking. Likewise, the AR feature is used as a straightforward visualization technique, guiding the surgeons without any external tool tracking but allowing them to do the final registration step between surgical tool and guidance system intuitively by themselves.

Calibration using the proposed multimodality marker is straightforward and proved to be reasonably accurate considering the current design. Use of the marker for calibration of the C-arm to the HMD is a convenient solution due to its flexibility that is slightly impeded by the need for prior offline calibration of the intrinsic parameters. Consequently, it would be beneficial to investigate other marker designs that would enable simultaneous calibration of intrinsic and extrinsic parameters of all cameras and thus promote the ease-of-use even further. In the same line of reasoning, although ARToolKit is a well-known tool for camera calibration via marker tracking, it might not be the best solution here as it is not designed for transmission imaging, which partly explains the low accuracy reported in the calibration experiment. The marker design itself, i.e., the sheet of lead, imposes the need for digital subtraction, which increases the required number of x-ray images and, thus, the dose by a factor of two, which may not be favorable in the clinical
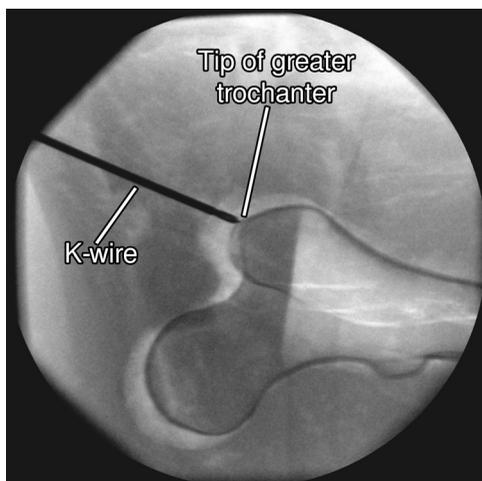
Fig. 6 X-ray of the second experiment on a semianthropomorphic femur phantom with the K-wire in the final position for one of the participants.

scenario. However, the need for subtraction is conditional on the design of the marker. Within this feasibility study, we valued convenient processing using ARToolKit over dose reduction and thus required subtraction imaging. This requirement, however, may become obsolete when transitioning to more advanced marker designs, e.g., by combining ARToolKit markers with small metal spheres for RGB and x-ray calibration, respectively. Such advanced approaches would further allow simultaneous calibration of both extrinsic and intrinsic parameters.

While the localization of the HMD with respect to the world coordinate system works well in most cases, it proved unreliable in scenarios where the surroundings are unknown, i.e., at the beginning of the procedure, or in presence of large changes in the environment, such as moving persons. While this shortcoming does affect the quantitative results reported here, it does not impair the relevance of the proposed guidance solution as more powerful devices and algorithms for SLAM will become available in the future.

The HMD used here adjusts the rendering of the virtual objects based on the interpupillary distance.[33] However, despite its name and advertisement, the HoloLens is not a holographic display as all virtual objects are rendered at a focal distance of ∼2 m. The depth cue of accommodation is thus not available. This is particularly problematic for unexperienced users, as the K-wire on the patient and the virtual objects designated for guidance cannot be perceived in focus at the same time. Moreover, there is currently no mechanism available that allows for a natural interaction between real and virtual objects. Consequently, cues that enable correct alignment of the tools with the guidance line, such as accurate occlusions or shading, cannot be provided. The difficulties in alignment are reflected in the staircase phantom of the feasibility studies. While these shortcomings affect the current prototype, we expect these challenges to be mitigated in the future when more sophisticated AR hardware becomes available.

4.3 Limitations

The proposed system required a wireless data sharing network to stream intraoperative images to the HMD. While this requirement may be considered a drawback at this very moment, it

may be seen as less unfavorable when intraoperative inspection of medical images transitions from traditional to virtual monitors.[15,34,35] Conceptually, it may even be possible to perform this "on-the-fly" guidance without any connection to the C-arm or other additional surgical hardware using only the HMD and the multimodality marker. In this version, the radiographic projection of the fiducial is directly observed on the physical monitor using the RGB camera of the HMD. Annotations can be made directly on an AR plane that overlies the radiographic image at the position of the physical monitor, potentially allowing for AR guidance in completely unprepared environments. However, use of the radiology monitor rather than the raw x-ray image introduces substantial additional sources of error into the system, as the pose of the HMD with respect to the monitor plane must be very accurately known. Furthermore, the ability to annotate the image with this method could be difficult depending on the position of the monitor in the room and the distance between the surgeon and the monitor.

Wearing an HMD may feel uncomfortable to some surgeons. We hypothesize that the need to wear an HMD during surgery is not a major impediment for orthopedic surgery, where head-based tools, such as magnification loupes, sterile surgical helmets, and headlamps, which are heavier than an HMD and many times require tethering, are already part of clinical routine.

5 Conclusion

We proposed an easy-to-use guidance system for orthopedic surgery that cocalibrates a C-arm system to an optical see-through HMD to enable on-the-fly AR in minimally prepared environments. Cocalibration of the devices is achieved using a multimodality marker that is then registered to the world coordinate frame using the SLAM tracking of the HMD. After calibration, point and line annotations in the 2-D x-ray images are rendered using the HMD as the corresponding virtual lines and planes in 3-D space, respectively, that serve as guidance to the surgeon.

The performance of the proposed system is promising for starting point localization in percutaneous procedures and could benefit from future advances of AR. Particularly, future work should consider possibilities to improve on the interaction of real and virtual objects, as the current lack of depth cues impedes superior performance.

Disclosures

The authors declare that they have no conflict of interest.

Acknowledgments

References

1. S. B. Gay et al., "Percutaneous screw fixation of acetabular fractures with ct guidance: preliminary results of a new technique," *AJR Am. J. Roentgenol.* **158**(4), 819–822 (1992).
2. B. M. Boszczyk et al., "Fluoroscopic radiation exposure of the kyphoplasty patient," *Eur. Spine J.* **15**(3), 347–355 (2006).
3. G. Kirousis et al., "Dosimetry during intramedullary nailing of the tibia patient and occupational exposure," *Acta Orthop.* **80**(5), 568–572 (2009).

4. I. D. Sugarman, I. Adam, and T. D. Bunker, "Radiation dosage during AO locking femoral nailing," *Injury* **19**(5), 336–338 (1988).

5. J. E. Robb, I. H. Annan, and M. F. Macnicol, "Guidewire damage during cannulated screw fixation for slipped capital femoral epiphysis," *J. Pediatr. Orthop. B* **12**(3), 219–221 (2003).

6. L. E. Zionts, P. Simonian, and J. Harvey Jr., "Transient penetration of the hip joint during in situ cannulated-screw fixation of slipped capital femoral epiphysis," *J. Bone Joint Surg.* **73**(7), 1054–1060 (1991).

7. F. Gras et al., "Screw placement for acetabular fractures: which navigation modality (2-dimensional vs. 3-dimensional) should be used? An experimental study," *J. Orthop. Trauma* **26**(8), 466–473 (2012).

8. S. Maqungo et al., "Distal interlocking screw placement in the femur: free-hand versus electromagnetic assisted technique (sureshot)," *J. Orthop. Trauma* **28**(12), e281–e283 (2014).

9. N. Hawi et al., "Radiological outcome and intraoperative evaluation of a computer-navigation system for femoral nailing: a retrospective cohort study," *Injury* **45**(10), 1632–1636 (2014).

10. A. Wilharm et al., "2D-fluoroscopic based navigation for Gamma 3 nail insertion versus conventional procedure: a feasibility study," *BMC Musculoskeletal Disord.* **14**(1), 74 (2013).

11. N. M. Hamming, M. J. Daly, and J. C. Irish, "Automatic image-to-world registration based on x-ray projections in cone-beam CT-guided interventions," *Med. Phys.* **36**, 1800–1812 (2009).

12. T. Doke et al., "Fluoroscopy-based laser guidance system for linear surgical tool insertion depth control," *Int. J. Comput. Assisted Radiol. Surg.* **10**, 275–283 (2015).

13. M. Kraus et al., "Computer-aided surgery does not increase the accuracy of dorsal pedicle screw placement in the thoracic and lumbar spine: a retrospective analysis of 2,003 pedicle screws in a level I trauma center," *Global Spine J.* **5**(2), 93–101 (2015).

14. J. Victor and D. Hoste, "Image-based computer-assisted total knee arthroplasty leads to lower variability in coronal alignment," *Clin. Orthop. Relat. Res.* **428**, 131–139 (2004).

15. L. Qian et al., "Towards virtual monitors for image guided interventions-real-time streaming to optical see-through head-mounted displays," arXiv:1710.00808 (2017).

16. N. Navab, S.-M. Heining, and J. Traub, "Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications," *IEEE Trans. Med. Imaging* **29**(7), 1412–1423 (2010).

17. J. Fotouhi et al., "Plan in 2D, execute in 3D: an augmented reality solution for cup placement in total hip arthroplasty," *J. Med. Imaging* **5**(2), 021205 (2018).

18. S. C. Lee et al., "Calibration of RGBD camera and cone-beam ct for 3D intra-operative mixed reality visualization," *Int. J. Comput. Assisted Radiol. Surg.* **11**(6), 967–975 (2016).

19. M. Fischer et al., "Preclinical usability study of multiple augmented reality concepts for K-wire placement," *Int. J. Comput. Assisted Radiol. Surg.* **11**(6), 1007–1014 (2016).

20. J. Fotouhi et al., "Pose-aware C-arm for automatic re-initialization of interventional 2D/3D image registration," *Int. J. Comput. Assisted Radiol. Surg.* **12**, 1221–1230 (2017).

21. E. Tucker et al., "Towards clinical translation of augmented orthopedic surgery: from pre-op CT to intra-op x-ray via RGBD sensing," *Proc. SPIE* accepted.

22. H. Wang et al., "Precision insertion of percutaneous sacroiliac screws using a novel augmented reality-based navigation system: a pilot study," *Int. Orthop.* **40**(9), 1941–1947 (2016).

23. H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Rob. Autom. Mag.* **13**(2), 99–110 (2006).

24. H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *2nd IEEE and ACM Int. Workshop on Augmented Reality (IWAR) Proc.*, pp. 85–94, IEEE (1999).

25. W. A. Chilcote et al., "Digital subtraction angiography of the carotid arteries: a comparative study in 100 patients," *Radiology* **139**(2), 287–295 (1981).

26. M. Unberath et al., "Virtual single-frame subtraction imaging," in *Proc. Int. Conf. Image Form X-Ray CT*, pp. 89–92 (2016).

27. J. Fotouhi et al., "Interventional 3D augmented reality for orthopedic and trauma surgery," in *16th Annual Meeting of the Int. Society for Computer Assisted Orthopedic Surgery (CAOS)* (2016).

28. J. Fotouhi et al., "Can real-time RGBD enhance intraoperative cone-beam CT?" *Int. J. Comput. Assisted Radiol. Surg.* **12**, 1211–1219 (2017).

29. L. Qian et al., "Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display," *Int. J. Comput. Assisted Radiol. Surg.* **12**, 901–910 (2017).

30. Microsoft, "Coordinate systems" (2017), https://developer.microsoft.com/en-us/windows/mixed-reality/coordinate_systems (13 August 2017).

31. B. C. Kress and W. J. Cummings, "11-1: invited paper: towards the ultimate mixed reality experience: HoloLens display architecture choices," *SID Symp. Dig. Tech. Pap.* **48**(1), 127–131 (2017).

32. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge (2003).

33. S. Hollister, "Microsoft HoloLens hands-on: incredible, amazing, prototype-y as hell" (2017), https://gizmodo.com/project-hololens-hands-on-incredible-amazing-prototy-1680934585 (13 August 2017).

34. P. C. Chimenti and D. J. Mitten, "Google Glass as an alternative to standard fluoroscopic visualization for percutaneous fixation of hand fractures: a pilot study," *Plast. Reconstr. Surg.* **136**(2), 328–330 (2015).

35. J. W. Yoon et al., "Technical feasibility and safety of an intraoperative head-up display device during spine instrumentation," *Int. J. Med. Rob. Comput. Assisted Surg.* **13**(3), e1770 (2016).

Sebastian Andress is an MD student at Ludwig Maximilian University in Munich since 2012 and conducts research with the trauma surgeons and computer scientists at the NARVIS Lab. His research interests include planning and guidance of surgeries, treatment of acetabular fractures, and 3-D printing.

Alex Johnson is a resident physician at the Department of Orthopaedic Surgery at Johns Hopkins. He is currently conducting research investigating advanced visualization methods for orthopedic surgical applications in collaboration with the Laboratory for Computer Aided Medical Procedures at Johns Hopkins University.

Mathias Unberath is a postdoctoral fellow at Johns Hopkins University. He received his BSc degree in physics, MSc degree in optical technologies, and PhD in computer science from Friedrich-Alexander-Universität Erlangen-Nürnberg in 2012, 2014, and 2017, respectively. Throughout 2014, he was appointed graduate research scholar at Stanford University. He is interested in medical image processing, and image reconstruction and visualization to improve diagnostic assessment and interventional guidance.

Alexander Felix Winkler is a graduate student of computer science at the Technical University of Munich, where he also earned his BSc degree in computer science in 2014, and currently a visiting student at Johns Hopkins University. He has been involved in several research projects at the NARVIS Lab as the Computer Aided Medical Procedures chair in Munich. His research interests include augmented and virtual reality, medical simulation, and 3-D printing.

Kevin Yu is a graduate student at the Laboratory for Computer Aided Medical Procedures at the Technical University of Munich and Johns Hopkins University. His research interests include RGB-D camera-based systems, communication between multiple camera-systems, and medical augmented reality.

Javad Fotouhi has been a PhD student at the Computer Science Department at Johns Hopkins University since 2014. Prior to joining the Laboratory for Computer Aided Medical Procedures at Johns Hopkins University, he earned his MSc degree in biomedical computing from the Technical University of Munich, and BSc degree in electrical engineering from the University of Tehran. His research focus includes x-ray interventional imaging, medical image registration and reconstruction, and augmented reality.

Simon Weidert has being an orthopedic trauma surgeon and researcher at the University of Munich, Germany, since he graduated from the University of Mainz in 2007. He works on computer-aided surgery with a focus on surgical simulation, augmented reality, and 3-D printing. He has cofounded two spin-offs from this field of research, Medability GmbH, and Mecuris GmbH.

Greg Osgood is an orthopedic trauma surgeon at Johns Hopkins Hospital. He graduated medical school at Columbia University College of Physicians and Surgeons and then completed residency training at New York Presbyterian Hospital. He completed his subspecialty orthopedic trauma fellowship at Harborview Medical Center. At Johns Hopkins, he is the first chief of orthopedic trauma in the hospitals history. His research focuses are advancing 2-D/3-D imaging, head-mounted displays, and advanced guidance for surgery.

Nassir Navab, MICCAI fellow, is a professor and a director of the Laboratory for Computer Aided Medical Procedures at the Technical University of Munich and Johns Hopkins University. He completed his PhD at INRIA and University of Paris XI in 1993. He received Siemens Inventor of the Year Award in 2001, SMIT Society Technology award in 2010, and "10 years lasting impact award" of IEEE ISMAR in 2015. His research interests include medical AR, computer vision, and machine learning.