

Development and Pre-Clinical Analysis of Spatiotemporal-Aware Augmented Reality in Orthopedic Interventions

Javad Fotouhi¹, Arian Mehrfard, Tianyu Song², Alex Johnson, Greg Osgood, Mathias Unberath³,
 Member, IEEE, Mehran Armand⁴, Member, IEEE, and Nassir Navab

Abstract—Suboptimal interaction with patient data and challenges in mastering 3D anatomy based on ill-posed 2D interventional images are essential concerns in image-guided therapies. Augmented reality (AR) has been introduced in the operating rooms in the last decade; however, in image-guided interventions, it has often only been considered as a visualization device improving traditional workflows. As a consequence, the technology is gaining minimum maturity that it requires to redefine new procedures, user interfaces, and interactions. The main contribution of this paper is to reveal how exemplary workflows are redefined by taking full advantage of head-mounted displays when entirely co-registered with the imaging system at all times. The awareness of the system from the geometric and physical characteristics of X-ray imaging allows the exploration of different human-machine interfaces. Our system achieved an error of 4.76 ± 2.91 mm for placing K-wire in a fracture management procedure, and yielded errors of $1.57 \pm 1.16^\circ$ and $1.46 \pm 1.00^\circ$ in the abduction and anteversion angles, respectively, for total hip arthroplasty (THA). We compared the results with the outcomes from baseline standard operative and non-immersive AR procedures, which had yielded errors of [4.61 mm, 4.76° , 4.77°] and [5.13 mm, 1.78° , 1.43°], respectively, for wire placement, and abduction and anteversion during THA. We hope that

our holistic approach towards improving the interface of surgery not only augments the surgeon's capabilities but also augments the surgical team's experience in carrying out an effective intervention with reduced complications and provide novel approaches of documenting procedures for training purposes.

Index Terms—Augmented reality, surgery, interaction, X-ray, frustum, visualization.

I. INTRODUCTION

INTERVENTIONAL image guidance is widely adopted across multiple disciplines of minimally-invasive and percutaneous therapies [1]–[4]. Despite its importance in providing anatomy-level updates, visualization of images and interaction with the intra-operative data are inefficient, thus requiring extensive experience to properly associate the content of the image with the patient anatomy. These challenges become evident in interventions that require the surgeon to navigate wires and catheters through critical structures under excessive radiation, such as in fracture or endovascular repairs.

Surgical navigation and robotic systems are developed to support surgery with localization and execution of well-defined tasks [5]–[8]. Though these systems increase the accuracy, their complex setup and explicit tracking nature may overburden the surgical workflow and consequently impede their acceptance in clinical routines [9]. Image-based navigation alleviates the requirements for external tracking, though depends strongly on pre-operative data which become outdated when the anatomy is altered during the surgery [10], [11].

As the expectations from surgical outcomes increase with more advanced technologies introduced in the ORs, the communication between the surgeon, crew, and the information becomes an important concern. Ineffective communication leads to increase of surgery time, radiation, and frustration to a point where in fluoroscopy-guided procedures, instead of the X-ray technician, the surgeons may reposition the scanners to ensure the task-defined views are optimal [12], [13].

To bridge the inefficiency gaps in surgical workflows, researchers have investigated the importance of human factor considerations in improving the usability of surgical data [14]–[16]. Recent works focused on facilitating the unmet interaction needs by introducing touch-less mechanisms such as gaze,

Manuscript received July 8, 2020; revised September 7, 2020; accepted November 1, 2020. Date of publication November 9, 2020; date of current version February 2, 2021. This work was supported in part by NIH under Award R01EB023939 and Award R01EB016703 and in part by Johns Hopkins University. (Javad Fotouhi, Arian Mehrfard, and Tianyu Song are co-first authors.) (Corresponding author: Javad Fotouhi.)

Javad Fotouhi, Tianyu Song, and Mathias Unberath are with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: javad.fotouhi@jhu.edu; tsong11@jhu.edu; unberath@jhu.edu).

Arian Mehrfard and Nassir Navab are with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD 21218 USA, and also with the Laboratory for Computer-Aided Medical Procedures, Technical University of Munich, 80333 Munich, Germany (e-mail: amehrf1@jhu.edu; nassir.navab@jhu.edu).

Alex Johnson and Greg Osgood are with the Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, MD 21218 USA (e-mail: ajohn190@jhmi.edu; gosgood2@jhmi.edu).

Mehran Armand is with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD 21218 USA, and also with the Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, MD 21218 USA (e-mail: marmand2@jhu.edu).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2020.3037013

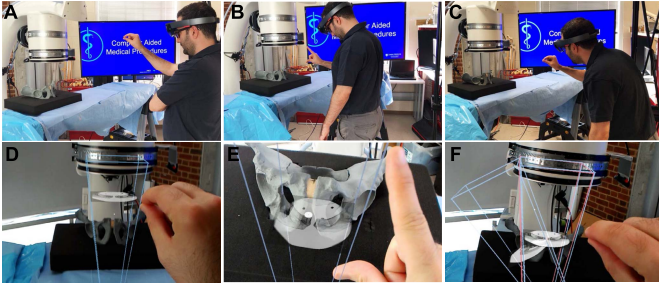


Fig. 1. Spatiotemporal-aware AR exploits the full imaging geometry. The augmented user interacts with the X-ray images within their viewing frustums (A-C). Corresponding AR views are shown in D-F.

foot, or voice commands [17], [18]. We believe the high stakes of surgery necessitates efficient interaction between all actors in the operating room *i.e.* surgeon, anesthesiologist and staff to communicate and access information. This demands user-centric designs that can also accommodate fluid movement of information and surgical inference across the entire team.

Augmented reality (AR) solutions have gained popularity in computer-integrated surgeries, as they can provide intuitive visualizations of medical data directly at patients' site. Early works on surgical AR focused largely on multi-modal fusion of information and provided display-based overlays [19]–[21]. Subsequently, AR enabled the utilization of pre- and intra-interventional 3D data during therapies [22]–[25].

The emergence of commercially available optical see-through head-mounted display (HMD) systems has led to development of AR solutions for various image-guided surgical disciplines, including percutaneous vertebroplasty, kyphoplasty, lumbar facet joint injection, orthopedic fracture management, bone cancer treatment, total hip arthroplasty (THA), interlocking nailing, cardiovascular surgeries, and surgical education [16], [26]–[33].

HMD-based AR has served as image viewer that directly displays the data at the operative site using virtual fluoroscopy monitors, hence eliminating the conventional off-axis visualization through static monitors [34]. Moreover, AR is used to provide navigational information during interventions [35]–[37]. These systems often rely on tracking of external markers, which require line-of-sight and invasive implantation into patients tissue that can hinder their usability. Andress *et al.* suggested a flexible marker-based surgical AR methodology which only required the marker to appear in the X-ray beam during the image acquisition, and was removed immediately after [38]. Recent inside-out localization strategies in AR have greatly favored the fluid workflow over explicit navigation, and have proved effective in eliminating the need for external markers [39]–[41].

This manuscript introduces the methodology and usability of a novel spatially-aware concept that enables immediate interaction with the medical data and promotes team approach where all stake-holders share a unified AR experience and communicate effectively. Our methodology exploits the viewing frustum of the imaging devices and human observers in the operating room (Fig. 1), and provides an engaging

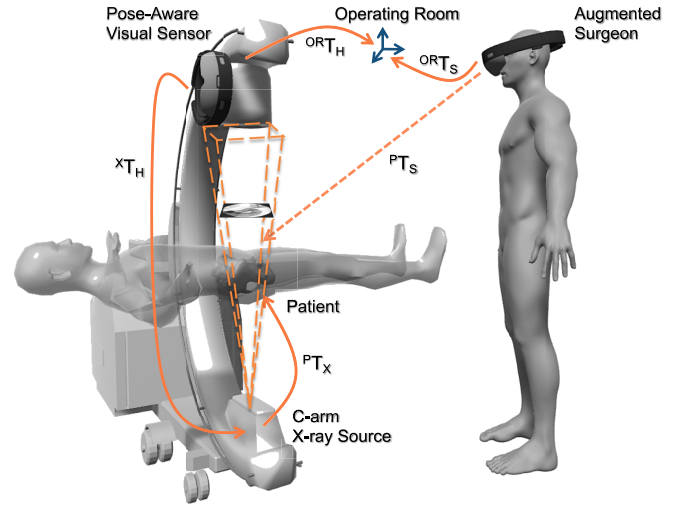


Fig. 2. The transformation chain of the spatially-aware AR system is shown for a C-arm fluoroscopy system. The transformations layout show the closed-loop between the imaging device and the users at all time. The same relations can be generalized to include multiple users.

and immersive experience for the surgical team. We built upon the concept of image frustums [41], and designed the complete methodology that integrates spatiotemporal-aware AR in the entire workflow. Our contributions are particularly centered around intra-operative planning, integration of the planning into the surgical workflow, and pre-clinical analysis. We showcase this solution in two high volume orthopedic procedures, *i.e.* K-wire placement in fracture care surgery, and acetabular cup placement in THA.

II. METHODOLOGY

Our main contributions are the collaborative AR concepts using spatiotemporal-aware flying frustums [41] that enable intra-operative planning, define new workflows, support surgical crew, enhance the communication between surgeon and data, and enable intuitive documentation of the surgery for training purposes. We present the methodology for the realization of these concepts in the remainder of this section.

A. Spatial-Awareness for AR

Visual data from cameras contain a wealth of information that can be used for simultaneous localization and mapping (SLAM). Visual SLAM is an important ingredient in our AR-based interaction recipe that enables co-localization of augmented users and the imaging device, which we will refer to as imaging observer, in a shared operating room environment. Marker-free co-localization will enable the 3D information to easily propagate through different bodies and be spatially-registered for all HMD users.

In the first step, each AR user is localized within the environment. The relative pose between two frames α and β using the environment map \mathbf{M} can be estimated by minimizing the following reprojection function:

$${}^{\alpha}\mathbf{T}_{\beta} = \arg \min_{{}^{\alpha}\hat{\mathbf{T}}_{\beta}} D({}^{\alpha}\hat{\mathbf{T}}_{\beta}, \mathbf{M})$$

$$\begin{aligned}
&= \arg \min \sum_{\substack{\alpha \hat{\mathbf{T}}_\beta \\ f_i^{(\alpha)} \in I_\alpha}} |f_i^{(\alpha)} - P(\mathbf{M}(f_i^{(\alpha)}))|^2 \\
&\quad + \sum_{f_i^{(\beta)} \in I_\beta} |f_i^{(\beta)} - P^\alpha \hat{\mathbf{T}}_\beta(\mathbf{M}(f_i^{(\beta)}))|^2, \quad (1)
\end{aligned}$$

where $f_i^{(\alpha)}$ and $f_i^{(\beta)}$ are corresponding features in images I_α and I_β , and P is the projection operator. In direct SLAM, the features include all pixels, and in indirect SLAM, the features are a sparse set of keypoints in the image. In Eq. 1, we optimize for the 6 parameters of a rigid transformation that best explains the pose between α and β using only the features present in the images. This step, also known as environment tracking, has become a standard part of most AR applications.

In a similar fashion, all users can be localized with respect to the first user, or with respect to a common spatial anchor in the operating room. The first member joining the shared experience will establish the anchor, *i.e.* OR coordinate system, and every other member of the AR session will share their pose in a master-slave configuration with respect to this OR frame [39]. This relation is shown as ${}^{\text{OR}}\mathbf{T}_S$ in Fig. 2.

B. Imaging Observer

C-arm scanners offer fluoroscopic imaging capabilities for a wide range of less-invasive therapeutic areas. To seamlessly integrate this imaging device into our interactive AR paradigm, we augment the scanner with a rigidly attached visual tracker, that observes the structures in the OR environment, and communicates spatial information to all users. The materialization of this imaging observer system requires a co-calibration between the visual tracker on the scanner, and the X-ray source [40]. The constant transformation that explains the calibration is denoted as ${}^{\text{X}}\mathbf{T}_H$ in Fig. 2.

To estimate ${}^{\text{X}}\mathbf{T}_H$, we formulate an over-determined system of equations as follows:

$$\begin{aligned}
{}^{\text{IR}}\mathbf{T}_{\text{OR}} &= {}^{\text{IR}}\mathbf{T}_{\text{X}}(t_i) {}^{\text{X}}\mathbf{T}_H {}^{\text{OR}}\mathbf{T}_H^{-1}(t_i) \\
&= {}^{\text{IR}}\mathbf{T}_{\text{X}}(t_{i+1}) {}^{\text{X}}\mathbf{T}_H {}^{\text{OR}}\mathbf{T}_H^{-1}(t_{i+1}). \quad (2)
\end{aligned}$$

IR denotes the frame of an external tracker that is used to track the motion of the C-arm source as it undergoes different motion at times t_i and t_{i+1} . To construct Eq. 2, the scanner is oriented in different poses, at each of which ${}^{\text{IR}}\mathbf{T}_{\text{X}}$ and ${}^{\text{OR}}\mathbf{T}_H$, which are the poses of the X-ray source in the external tracker frame and the SLAM-based localization of the HMD in the OR coordinate system, respectively, are recorded. It is important to note that the IR tracker is only used for this one-time and offline co-calibration, and it is not used intra-operatively. By re-arranging Eq. 2, we formulate the problem in the form of $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B}$ as presented in Eq. 3, such that $\mathbf{X} := {}^{\text{X}}\mathbf{T}_H$, and \mathbf{A} and \mathbf{B} represent the relative motion of the X-ray source and the SLAM capable visual sensor on the gantry, respectively.

$${}^{\text{IR}}\mathbf{T}_{\text{X}}^{-1}(t_{i+1}) {}^{\text{IR}}\mathbf{T}_{\text{X}}(t_i) {}^{\text{X}}\mathbf{T}_H = {}^{\text{X}}\mathbf{T}_H {}^{\text{OR}}\mathbf{T}_H^{-1}(t_{i-1}) {}^{\text{IR}}\mathbf{T}_{\text{X}}^{-1}(t_i). \quad (3)$$

Rotation and translation components of the hand-eye problem are disentangled and computed separately as:

$$R_A R_X = R_X R_B$$

$$R_A \mathbf{t}_X + \mathbf{t}_A = R_X \mathbf{t}_B + \mathbf{t}_X. \quad (4)$$

We estimate the rotation parameters using unit quaternion representation as $q_A q_X = q_X q_B$. Given that a unit quaternion q_i is formed by a vector \mathbf{v}_i and a scalar s_i such that $q_X = \mathbf{v}_X + s_X$, we re-write the rotation component in Eq. 4 using the quaternion product rule as:

$$\begin{aligned}
(\cdot) : s_A \mathbf{v}_X + s_X \mathbf{v}_A + \mathbf{v}_A \times \mathbf{v}_X &= s_X \mathbf{v}_B + s_B \mathbf{v}_X + \mathbf{v}_X \times \mathbf{v}_B \\
(\cdot) : s_A s_X - \mathbf{v}_A \cdot \mathbf{v}_X &= s_X s_B - \mathbf{v}_X \cdot \mathbf{v}_B. \quad (5)
\end{aligned}$$

Re-arranging the above formulation yields:

$$\begin{bmatrix} s_A - s_B & (\mathbf{v}_A - \mathbf{v}_B)^\top \\ (\mathbf{v}_A - \mathbf{v}_B) & (s_A - s_B)I_3 + [\mathbf{v}_A + \mathbf{v}_B]_\times \end{bmatrix} \begin{bmatrix} s_X \\ \mathbf{v}_X \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{0}_3 \end{bmatrix}, \quad (6)$$

which is then solved in a constrained optimization fashion as:

$$\min \|M\mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{q}\|_2^2 = 1, \quad (7)$$

where $\mathbf{q} = [s_X, \bar{\mathbf{v}}_X]^\top$. After the rotation parameters are computed, the translation vector is estimated in a least-squares setting: $(R_A - I_3)\mathbf{t}_X = R_X \mathbf{t}_B - \mathbf{t}_A$.

At the end of this phase, we have achieved R_X and \mathbf{t}_X that express the calibration between the X-ray source and the integrated visual tracker; hence, we directly access the pose of the X-ray camera from the pose information acquired from the visual tracker on the scanner.

C. Geometry-Awareness for AR

In this section, we describe the underlying geometry that allows us to combine the content of 2D X-ray images, directly with the 3D spatial information we computed in Sec. II-A and II-B. To this end, we explicitly model the viewable region of the X-ray camera, known as the flying frustum [41], and allow interaction with images within their geometries. It is important to note that, the flying frustum refers to the full pyramid of vision (Fig. 2), and is different than the truncated pyramids used in the computer graphics community. Despite the similarities in formulation, the conventional frustum model in graphics only applies to reflective images, and cannot accommodate the transmission model used in fluoroscopy. Therefore, we extend the perspective pinhole camera model that is commonly used in the computer vision community [42].

In our paradigm, users can move the images within their frustums on a virtual plane known as the *near* plane, between X-ray source and detector (referred to as the *far* plane), while they remain a valid image of the same anatomy. This interaction enables the users to intersect the images with corresponding anatomies, and intuitively observe *2D-image-to-3D-anatomy* associations. Additionally, the imaging technologists who operate the scanner, can align the scanner with a desired frustum that is decided by the surgeon.

A flying frustum is defined using the following model:

$$P_f = \begin{bmatrix} \frac{n}{f} & 0 & 0 \\ 0 & \frac{n}{f} & 0 \\ 0 & 0 & 1 \end{bmatrix} K P \begin{bmatrix} {}^{\text{OR}}R_X & {}^{\text{OR}}\mathbf{t}_X \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (8)$$

where n refers to the distance to the *near* plane, f is the focal length, K is the matrix of intrinsic parameters, and $0 \leq n \leq f$.

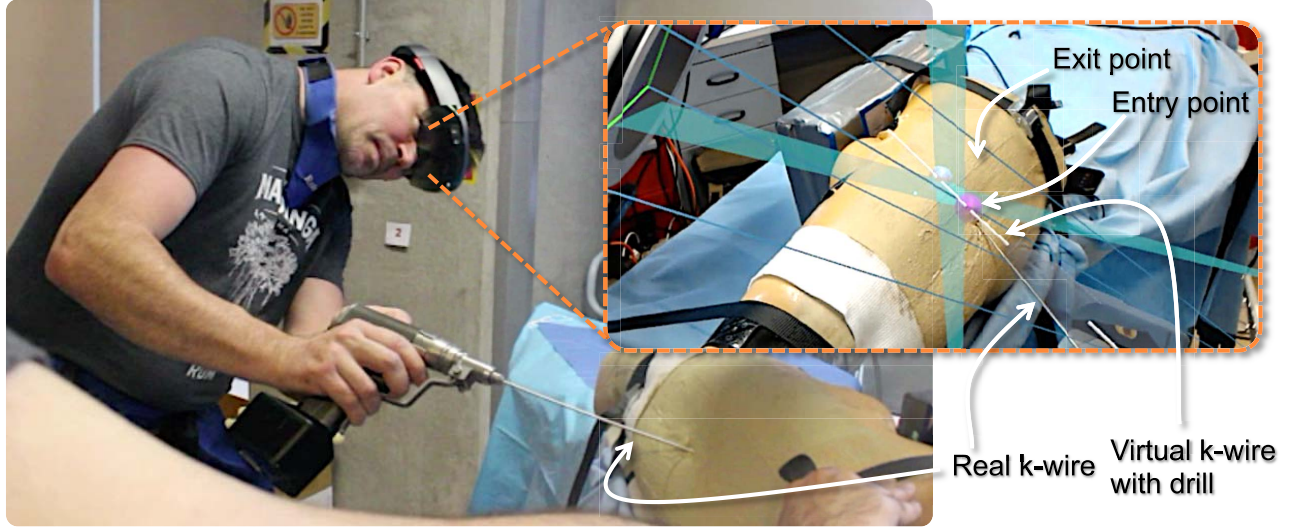


Fig. 3. In K-wire placement for fracture reduction procedures, the surgeon can plan the entry and exit points of the wire on two X-ray images. After the planning, two triangular planes are constructed by connecting the drilling trajectory defined on the detector plane (X-ray image) and the C-arm source (X-ray origin). The intersection of these two planes is a line that corresponds to the desired drilling trajectory in 3D. By exploiting the imaging frustum, this line is augmented directly on the patient anatomy. The surgeon can then align the physical drill with its virtual counterpart, and advance the wire through the anatomy.

The parameter n is controlled by the user, such that when $n = f$, the X-ray image is directly displayed at the detector scale. It is worth mentioning that, with conventional frustum models, the *near* plane can only take values smaller than the *far* plane, which is not the case in our representation.

For each 2D point $x_i \in \mathcal{I}$, where \mathcal{I} is the domain of all acquired images, the corresponding point x_f in the frustum domain \mathcal{F} is scaled by a factor s as $\mathbf{x}_f = s \mathbf{x}_i = (\frac{n}{f})\mathbf{x}_i$, such that $0 \leq n \leq f$. Finally, the 3D pose of the interactive image in the frustum is defined as:

$$\begin{aligned} {}^{\text{OR}}\mathbf{T}_1 &= \begin{bmatrix} {}^{\text{OR}}R_X & {}^{\text{OR}}\mathbf{t}_X \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} I_3 & 0 \\ \mathbf{0}^\top & n \\ \mathbf{0}^\top & 1 \end{bmatrix} \\ &= \begin{bmatrix} {}^{\text{OR}}R_X & r_{13}n \\ & r_{23}n + {}^{\text{OR}}\mathbf{t}_X \\ \mathbf{0}^\top & r_{33}n \\ & 1 \end{bmatrix}, \end{aligned} \quad (9)$$

where $R = \{r_{i,j}\}_{i,j:1,2,3}$.

D. Planning Using Flying Frustums

Flying frustums discussed in Sec. II-A-II-C embed sufficient 3D and 2D information that enable interventional planning for the placement of surgical tools (Fig. 3). In this section we introduce two distinct approaches for intra-operative planning. In the first planning approach, we forward-project the 3D virtual surgical implants onto the X-ray images within each frustum using the respective X-ray projective geometry. Hence, the user can observe the implant's projection in the same X-ray image and verify the implant's appearance before placing the real implant. This approach is generic and allows the planning of implants with any arbitrary shape. In contrast to the earlier

works that overlay on the X-ray image the surgical tools that are tracked using navigation systems [43], our methodology enables intra-operative planning by projecting and moving a virtual implant (not a physically tracked implant) until it aligns with target structures in all valid frustums and X-ray images.

In the second approach, we use multi-view geometry to reconstruct landmark targets or trajectories in 3D. In both methods, after the respective planning on the flying frustums, the resulting 3D information, already registered to the anatomy, is visualized on the patient. In the following, we describe each of the two approaches.

In the first method, virtual tools are manipulated in 3D by the user, and simultaneously projected onto the X-ray images of all valid frustums. A point $X_t \in \mathcal{T}$, where \mathcal{T} is the domain of all 3D points on a virtual tool, is projected onto the i^{th} frustum as $x_{ti} = P_{fi}X_t$. The virtual tool is manipulated with complete 6 *degrees-of-freedom* (DOF) by the user to plan on all the frustums simultaneously. However, it may appear challenging to align concurrently in both views. Alternatively, we can apply rotational and translational constraints to the virtual tool, such that the tool can be aligned first in one frustum, and then in the second frustum, while the alignment in the first one is preserved. As shown in Fig. 4, in the first stage, the Y-axis of the virtual tool is rotated to hold the same direction as the Z-axis of the first frustum. The 4 DOF transformation model of the virtual tool is defined as:

$$\mathbf{T}^{f1} = \begin{bmatrix} \cos \theta & 0 & \sin \theta & t_x \\ 0 & 1 & 0 & t_y \\ -\sin \theta & 0 & \cos \theta & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

where θ is the rotation angle around the local Y-axis and t_x , t_y and t_z are translations along X, Y and Z-axes, respectively.

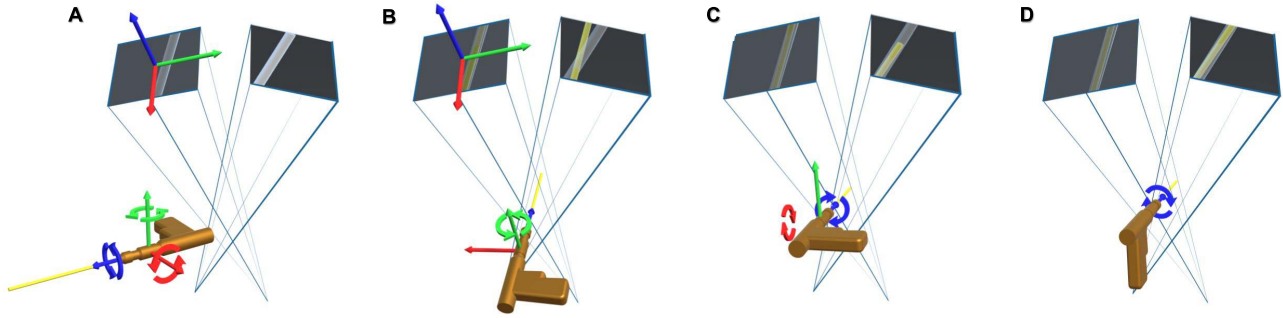


Fig. 4. The coordinate frame of the virtual drill is defined as shown in **A**, where the Z-axis that is shown in blue points out of the drill along with the K-wire. The corresponding X- and Y-axes are shown in red and green, respectively. **B** shows the manipulation of the virtual drill with 4 DOF until the projection of the K-wire is aligned in the first frustum (the yellow projection of the wire is contained within the tube in the left frustum). These 4 DOFs are sufficient to align the tool appropriately with the target anatomy in the first image. In the next step, we change the transformation constraints, as shown in **C**; after the alignment of the drill with the anatomy is verified in the first frustum, the drill maintains rotational freedom around its local X- and Z-axes, and translational freedom along its Y- and Z-axes. These DOF constraints allow the implant's alignment in the second frustum while maintaining the alignment between tool and anatomy in the first frustum. Finally, **D** shows the virtual drill being restricted to only 2 DOF. Moving and rotating along these two DOFs will not influence the alignment in either of the two frustums.

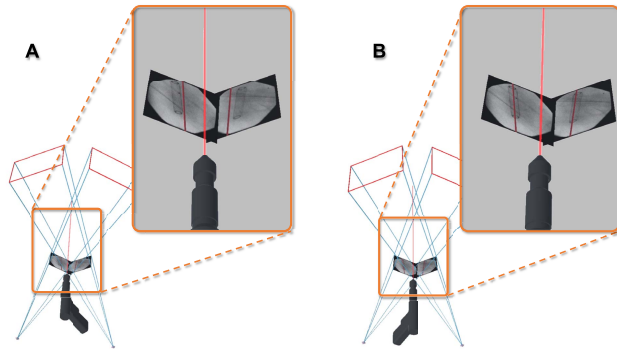


Fig. 5. The augmented projections allow us to exploit the geometry in AR and plan surgical tools in relation to patient anatomy. The misaligned virtual drill in **A** is repositioned until it appears inside the desired structure in all the frustums (**B**).

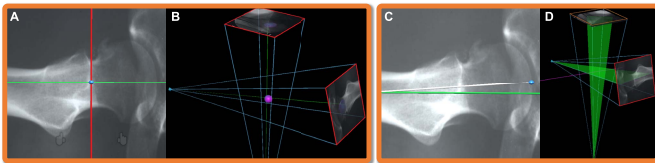


Fig. 6. Each point in a frustum image corresponds to a ray passing through the landmark in 3D, and connecting the source and detector of the C-arm. Intersection of two rays recovers the 3D point and renders it directly on the patient (**A-B**). Similarly, annotation of lines in each frustum, corresponds to a plane in 3D. The intersection of these planes restores the 3D planning trajectory, and renders it in AR such that it travels through the corresponding anatomical structure (**C-D**).

In the second stage, the virtual tool is locked to only allow rotation around the X and Z-axes, and translation in Y and Z-axes using the transformation model as:

$$\mathbf{T}^{f2} = \begin{bmatrix} \cos \psi & -\sin \psi \cos \phi & -\sin \psi \sin \phi & 0 \\ \sin \psi & \cos \psi \cos \phi & -\cos \psi \sin \phi & t_y \\ 0 & \sin \phi & \cos \phi & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

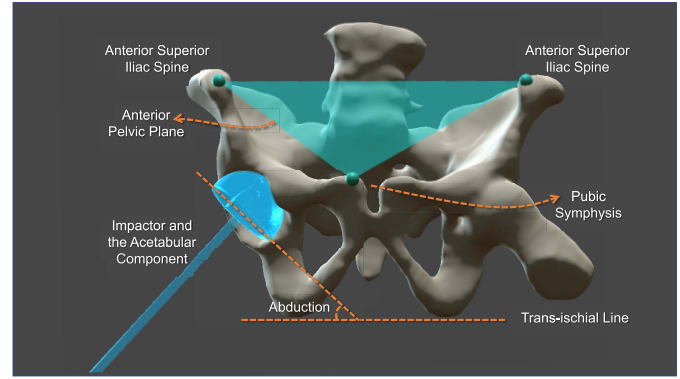


Fig. 7. In THA, abduction and anteversion angles of the acetabular implant are defined with respect to the anterior pelvic plane (APP). Abduction refers to the in-plane rotation, and anteversion refers to the out-of-plane rotation of the cup. The anterior pelvic plane is defined based on three points: the left and right anterior superior iliac spine landmarks, and the pubic symphysis. Once the surgeon annotates these landmarks, we identify the APP, and subsequently, render the acetabular components at appropriate angles. We also let the user annotate the center of the acetabulum in two or more views, which is used to calculate the 3D position of this landmark on the patient, hence allowing the center of the hemispheric component to render inside the hip socket. In the execution phase, the user aligns the real impactor and cup with their virtual counterparts.

where ϕ and ψ are the rotation angles around their X and Z-axes, respectively. As shown in **Fig. 4**, \mathbf{T}^{f2} does not influence the alignment of the tool and target in the image of the first frustum. Finally, after alignment in both frustums is achieved, the virtual tool then will only be constrained to have 2 DOF which is defined as:

$$\mathbf{T}^{tool} = \begin{bmatrix} \cos \psi & -\sin \psi & 0 & 0 \\ \sin \psi & \cos \psi & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (12)$$

In an exemplary case shown in **Fig. 5**, the virtual drill is rotated and translated until it passes through a desired structure (e.g. through a bone canal) in all frustums. An alignment

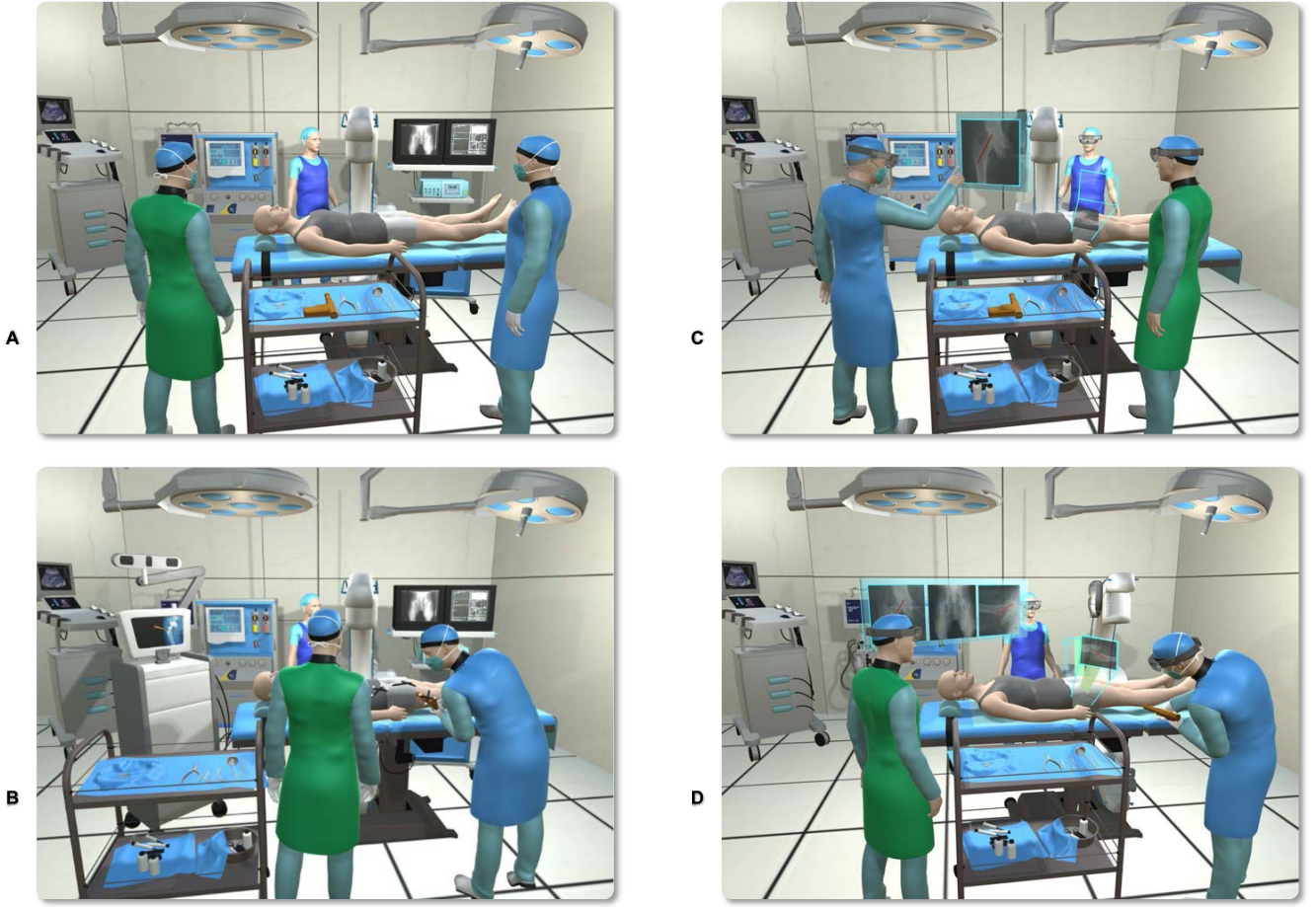


Fig. 8. The standard operative procedure in percutaneous orthopedic interventions makes extensive use of interventional imaging (A). Classic navigation-based solutions use sophisticated tracking hardware and external markers to provide geometric registration between the content in the image and the patient (B). On the other hand, in the AR-enhanced OR that we suggest (C), the surgeon and crew interactively use the data and pass the information around without explicit navigation. Based on the concepts introduced in Sec. II-D, the planning on X-ray images is directly visualized on the patient. The surgeon takes action based on the information from planning, as well as the X-ray images that are positioned within their respective frustums, both of which are seen through the HMD (D).

consensus in all frustums is the equivalent of the alignment of the virtual 3D tool with the imaged anatomy in 3D.

The second planning approach requires 2D interaction on the frustum X-ray images. In this setting, for each selected landmark on a frustum image, a 3D ray connecting the C-arm source and the target landmark will be rendered into the AR scene. As illustrated in Fig. 6, the intersection of two rays from a corresponding landmark in two images reconstruct the 3D landmark. Each ray is defined via two elements: *i*) the position of the C-arm X-ray source \mathbf{c}_i , and *ii*) the unit direction vector \mathbf{u}_i from the source to the annotated landmark in the frustum. We estimate the closest point \mathbf{x}_l^* to the $N = 2$ rays corresponding to each landmark l via a least-squares minimization strategy as follows:

$$\mathbf{x}_l^* = \arg \min_{\mathbf{x} \in \mathbb{R}^3} \sum_{i=1}^N \|(I_3 - \mathbf{u}_i \mathbf{u}_i^T) \mathbf{x} - \mathbf{t}_i\|^2, \quad \text{where } \mathbf{t}_i = (I_3 - \mathbf{u}_i \mathbf{u}_i^T) \mathbf{c}_i. \quad (13)$$

Similarly, two points on a frustum i defining the entry and the exit points of a drilling trajectory, associate to two rays \mathbf{u}_{1i}

and \mathbf{u}_{2i} in 3D. These two rays span a plane in 3D as shown in Fig. 6. The intersection of the planes corresponding to the same entry and exit points on frustums i and j form a 3D line $\mathbf{d}_{12} = (\mathbf{u}_{1i} \times \mathbf{u}_{2i}) \times (\mathbf{u}_{1j} \times \mathbf{u}_{2j})$ that passes through the desired entry and exit points on the patient anatomy.

Our first approach requires a more complex interaction with the augmented surgical implant using the 6 *degrees-of-freedom*, however generalizes to arbitrary structures beyond linear annotations, such as the curved plates used for internal fixations. In our second approach, if the 3D model of the implant is known, the planning can be achieved by annotating only three corresponding landmarks in each frustum image.

To perform intra-operative planning with flying frustums, two X-ray images are sufficient. During these two acquisitions, we assumed that the patient is static. For many orthopedic and trauma interventions, particularly for K-wire placement, this is an entirely valid assumption.

E. Surgical Workflow Integration

Intra-operative planning and execution with the flying frustums support can be used in various fluoroscopy-guided



Fig. 9. All acquisitions can be documented and later reviewed with all their corresponding spatial and temporal information. Spatiotemporal-aware AR allows the trainees to watch the surgery's progress and revisit the actions taken based upon each image.

procedures. In THA, the critical points defining the anterior pelvic plane (APP) can be each identified on X-ray images. These anatomical landmarks include the left and right anterior superior iliac spine points on the pelvic wing and the pubic symphysis. Given APP, a virtual acetabular implant and a rigidly attached impactor are rendered in AR with their desired orientation that is calculated with respect to APP. Likewise, the translational component of the cup implant is identified by defining the center of the patient acetabulum on corresponding fluoroscopic images. These relations are shown in Fig. 7. Once these intra-operative planning steps are completed by the surgeon, virtual representations of the cup and impactor are augmented over the patient's acetabulum with the appropriate abduction and anteversion angles. The surgeon could then align the impactor with its virtual counterpart reducing the amount of required X-rays. It is important to note that achieving the desired angles for the hip implant is a crucial step that is mentally challenging to verify from single view X-ray images, as it is commonly practiced in the direct anterior approach.

Another exemplary image-guided procedure is the placement of screws and K-wires during fracture management. As shown in Fig. 8, AR provides support for placement of K-wires using the trajectory planning on the corresponding frustums. Fig. 8 also depicts the use of our AR solution in the OR, and compares the proposed environment with the current OR.

Our proposed AR landscape is enabled by exploiting all involved frustums to move spatial information between different bodies, hence allowing multiple users to connect simultaneously. In Fig. 8, we show hypothetical procedures and further demonstrate that the scanner, crew and the technician can all share this common AR experience through HMDs, thus jointly benefiting from the augmented procedure. As highlighted in the figure, our system relies on 2D C-arm fluoroscopy, thus the standard workflow is only minimally altered.

The surgeon can always alternate between fluoroscopy-based guidance and the AR view to ensure safe drilling. The figure also signifies the advantage that the surgeon does not need to take his/her gaze away from the patient site during implant placement. As shown in Fig. 9, all the spatial and temporal information can be documented for post-operative review and training.

III. EXPERIMENTAL RESULTS

A. System Setup

Our system comprises an ARCADIS Orbic 3D C-arm (Siemens Healthineers, Forchheim, Germany) as an intra-operative X-ray device that automatically computes the cumulative area dose for each session. The immersive AR solution was built using the Unity cross-platform game engine (Unity Technologies, San Francisco, CA, US) and was deployed to an optical see-through HMD, the Microsoft HoloLens (Microsoft, Redmond, WA). To jointly co-localize the augmented surgeon and the C-arm scanner, a second HoloLens device with inside-out SLAM capabilities was attached near the X-ray detector. The two HMDs shared their spatial anchor, a rich feature reference region in the common environment, over a wireless local network, allowing them to remain synchronized and establish spatial awareness. This connection was enabled through a TCP-based sharing service running on an Alienware (Dell, Round Rock, TX, US) laptop server with an Intel i7-7700HQ CPU, NVIDIA GTX 1070 graphics card, 16 GB RAM, and Windows 10 operating system. The 16 bit 1024×1024 single channel X-ray images from C-arm were transmitted to the server computer over a direct Ethernet connection, and then converted to 8 bit grayscale while keeping the original resolution in order to display them properly on the HMD as well as to reduce the data size before uploading them to the HMD. The frame per second update for the HMD device was 60, and the display resolution was 720p.

To minimize the chance of tracking failure, we mounted the tracking HMD on the C-arm such that the integrated vision cameras on the HMD observe the static structures such as walls and ceiling. Consequently, the chances of blocking the tracking cameras on the HMD during the procedure are diminished.

B. System Calibration

To solve the hand-eye calibration problem in Eq. 3, 120 different pairs of corresponding poses were recorded from the visual tracker on the C-arm as well as an external infrared tracking system that tracked the C-arm source. Fig. 10 presents the error for this offline calibration step given different sampling for the pose pairs.

The localization quality of the SLAM-based visual tracking system on the scanner, *i.e.* the HMD on the scanner, was compared to a ground-truth provided by an external tracker. We measured rotational errors of $(0.71^\circ, 0.11^\circ, 0.74^\circ)$ with a norm of 0.75° and translational errors of $(4.0, 5.0, 4.8)$ mm with a norm of 8.0 mm along the (X, Y, Z) -axes, respectively [41].

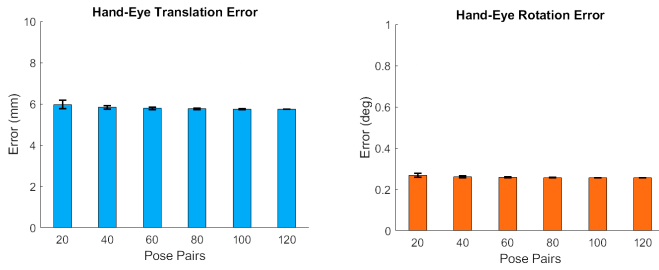


Fig. 10. Mean and standard deviation of the translational and rotational errors for the hand eye calibration step are shown in the left and right plots, respectively. For each N number of pose pairs shown on the horizontal axis (except the last column which considers all the available data), the experiments were repeated 100 times by each time sampling N poses from the total 120 available poses and computing the hand-eye calibration.

C. Experiments

Eight orthopedic surgeons and residents from Johns Hopkins Hospital participated in pre-clinical user studies and performed two surgically relevant tasks while utilizing interactive flying frustums in an immersive AR environment. The surgeons are referenced as **P1**. **P6** is an attending surgeon. At the time of the study, **P7** and **P4** were in their final year of residency, **P5** in their fourth year, **P3** in their third year, **P2** in their second year, and **P1** and **P8** were first-year residents.

In the first procedure, we focused on the correct placement of a K-wire to repair complex fractures. To emulate the K-wire placement through the superior pubic ramus (acetabulum arc), we used radiopaque cubic phantoms, as seen in **Fig. 11-C**. For direct comparison, we used the same setup that was used by Fischer *et al.* [22]. Each cube consisted of a stiff, lightweight, and non radiopaque methylene bisphenyl diisocyanate (MDI) foam. Since the superior pubic ramus is a tubular bone with a diameter of approximately 10 mm, we used a thin aluminium mesh filled with MDI that was placed inside each cube and served as the bone phantom. The two ends of the tubular structures were complemented with a rubber radiopaque ring. Each subject was asked to place a K-wire with a diameter of 2.8 mm through the tubular phantom using a surgical drill (Stryker Corporation, Kalamazoo, MI, US).

For the second procedure, we constructed a total hip arthroplasty mock setup by using a radiopaque pelvis phantom with a magnetic acetabulum to fixate the acetabular cup (**Fig. 12**). For direct comparison, we adopted the same experimental setup that was suggested by Alexander *et al.* [44]. The cup was attached to a straight cylindrical acetabular trialing impactor (Smith & Nephew, London, UK) allowing the operator to guide the cup. Since the ideal orientation of the implant is unknown, we use abduction and anteversion angles that lie in a safe zone defined by landmarks on the pelvis as described in [45].

Initially, each surgeon received a brief introduction to the Microsoft HoloLens, preparing them to properly mount and use the HMD. To further instruct them on our AR application, pre-recorded training X-ray images were loaded onto their HMD, allowing them to become familiar with the interface, planning procedure, and the interaction mechanism using hand gestures.

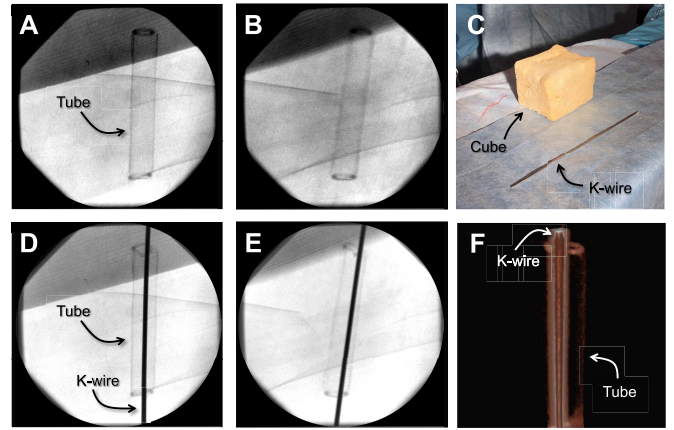


Fig. 11. **A-B** are the X-ray images of the cubic phantom shown in **C**. In **D-E**, the X-ray images of the same phantom are shown after a K-wire was successfully inserted inside the tube. **F** is the CBCT scan of the phantom which was acquired for verification. Due to metal artifacts, the tube does not exhibit strong contrast.

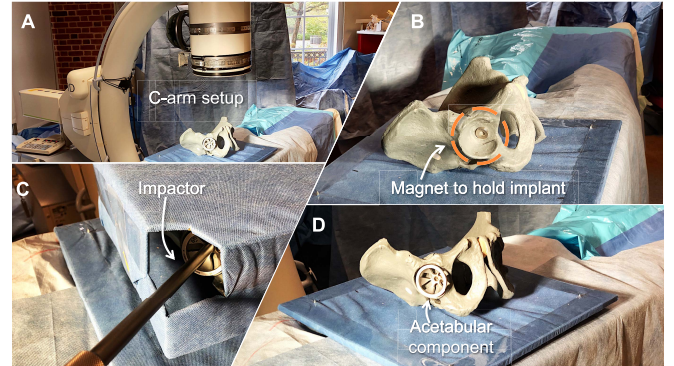


Fig. 12. In **A** the setup of the C-arm, pelvic phantom, and the acetabular cup are shown. **B** is a close-up view of the phantom with an empty acetabular socket and a magnet for holding the implant in position. Image **C** shows the impactor while it is placed by a surgeon during the experiment, and **D** shows the successfully placed cup in the acetabulum.

After the required planning images were acquired by the proctors, each surgeon planned their respective procedure in AR and performed the drilling task into the cube or placed the acetabular component into the pelvis. During the procedure, they were explicitly allowed to order as many X-ray shots from any perspective that they considered necessary.

We recorded the planing time, the time it took them to execute the procedure, number of fluoroscopic acquisitions, and the cumulative radiation dose as it was measured by the scanner. Finally, for the verification and accuracy measurement, we acquired a 3D cone-beam CT (CBCT) scan of the phantoms with their respective implants. It is important to note that distance is not defined for two non-parallel lines. The distance we computed from the drilled path to the desired path, i.e., the average distance from the wire to the center-line of the target structure, is consistent with the past literature [22], and provides an intuition regarding the range of error.

D. Results

Tables I and **II** comprise the performance of every participant in the experiments. **Table I** contains the measurements

TABLE I

OUTCOME FROM THE K-WIRE INSERTION USING OUR IMMERSIVE AR SYSTEM. INDIVIDUAL PERFORMANCES ARE LISTED IN COLUMNS P_i. CORRESPONDING MEAN AND SD VALUES CAN BE FOUND IN TABLE IV. THE LAST ROW REPORTS THE ERROR THAT WAS MEASURED BETWEEN THE INSERTED K-WIRE AND THE CENTER OF THE TUBE

K-wire	P1	P2	P3	P4	P5	P6	P7	P8
Planning Time (sec)	125	40	39	58	79	22	67	44
Execution Time (sec)	83	74	58	46	66	67	63	79
# X-ray images	2	2	2	2	2	2	2	2
Dose (cGY(cm ²))	0.28	0.21	0.19	0.27	0.26	0.28	0.27	0.28
Error (mm)	8.23	5.71	9.02	3.26	6.94	1.13	1.59	2.23

TABLE II

OUTCOME FROM THE PLACEMENT OF THE ACETABULAR IMPLANT USING OUR IMMERSIVE AR SYSTEM. INDIVIDUAL PERFORMANCES ARE LISTED IN COLUMNS P_i. CORRESPONDING MEAN AND SD VALUES CAN BE FOUND IN TABLE V

THA	P1	P2	P3	P4	P5	P6	P7	P8
Planning Time (sec)	162	70	117	88	64	37	71	110
Execution Time (sec)	87	39	13	19	17	35	26	24
# X-ray images	8	8	8	8	8	8	8	8
Dose (cGY(cm ²))	1.27	1.3	1.23	1.26	1.18	1.25	1.18	1.29
Abduction error (°)	2.1	1	1.1	1.3	2.9	0.1	0.4	3.7
Anteversión error (°)	1.1	0.6	2.7	1.4	2.1	0.4	0.3	3.1

TABLE III

RESULTS OF THE RESPECTIVE SOP PRESENTED IN [22] AND [44]. COLUMNS CORRESPOND TO INDIVIDUAL PARTICIPANTS PERFORMANCE. CORRESPONDING MEAN AND SD VALUES CAN BE FOUND IN TABLE IV AND TABLE V

K-wire SOP							
	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Time (sec)	937	686	617	464	636	388	432
# X-ray images	80	47	44	33	32	21	29
Dose (cGY(cm ²))	7.68	1.73	3.54	4.38	5.62	2.69	5.38
Error (mm)	3.08	7.88	11.43	3.01	1.87	2.27	2.72

THA SOP								
	R1	R2	R3	R4	R5	R6	R7	R8
Time (sec)	210	350	195	375	225	150	95	280
# X-ray images	13	15	15	16	13	13	6	19
Dose (cGY(cm ²))	1.93	2	2.09	2.11	1.83	1.88	1.09	2.73
Abduction error (°)	7.9	5.9	5.1	6.5	5.6	2.3	1.5	3.3
Anteversión error (°)	4.9	1.3	9.9	5.1	1.4	6	1.8	7.8

TABLE IV

MEAN AND SD VALUES FOR K-WIRE INSERTION WITH THE IMMERSIVE AR, NI-AR, AND SOP. FOR EACH METHOD TWO ROWS SHOW THE MEAN AND SD VALUES, RESPECTIVELY. FOR IMMERSIVE AR, THE TIME IS SEPARATED INTO FIRST PLANNING AND THEN EXECUTION

Method	Time (sec)	# X-ray images	Dose (cGY(cm ²))	Error (mm)
\bar{X} AR	59.25 + 52	2	0.255	4.76
σ	(32.02, 24.23)	(0)	(0.04)	(3.11)
\bar{X} NI-AR[22]	243.7	2.14	1.6	5.13
σ	(84.00)	(0.69)	(0.17)	(2.72)
\bar{X} SOP	594.3	40.86	4.43	4.61
σ	(188.0)	(19.38)	(2.00)	(3.62)

TABLE V

MEAN AND SD VALUES FOR ACETABULAR CUP PLACEMENT WITH THE IMMERSIVE AR, NI-AR, AND SOP. FOR EACH METHOD TWO ROWS SHOW THE MEAN AND SD VALUES, RESPECTIVELY. FOR IMMERSIVE AR, THE TIME IS SEPARATED INTO FIRST PLANNING AND THEN EXECUTION TIME. IN THE #X-RAY COLUMN OF NI-AR, ONLY ONE X-RAY IS DENOTED, THIS REFERENCES THE CBCT THAT WAS ACQUIRED BEFORE THE EXPERIMENT WHICH IS RECONSTRUCTED OF 100 DIGITAL RADIOGRAPHS

Method	Time (sec)	# X-ray images	Dose (cGY(cm ²))	Abd. (°)	Ant. (°)
\bar{X} AR	89.88 + 32.5	8	1.25	1.57	1.46
σ	(38.85, 23.71)	(0)	(1.25)	(1.24)	(1.07)
\bar{X} NI-AR [44]	110.6	1 CBCT	1.83	1.78	1.43
σ	(15)	(0)	(0.06)	(1.37)	(0.66)
\bar{X} SOP	235	13.75	1.96	4.76	4.78
σ	(96)	(3.73)	(0.45)	(2.2)	(3.15)

for the K-wire insertion, and Table II presents the procedural outcome for the acetabular cup placement. We separate the interventional time measurements into i) *planning time*, the time it took each surgeon to plan their procedure in AR, and ii) *execution time*, determining the duration of the insertion/placement of the instruments. Furthermore, we recorded

the number of X-ray acquisitions and the respective dose for each user. Finally, to assess the overall performances, we computed the average distance of the K-wire from the center of the tube at the entry and the exit surface of the tubular structure,

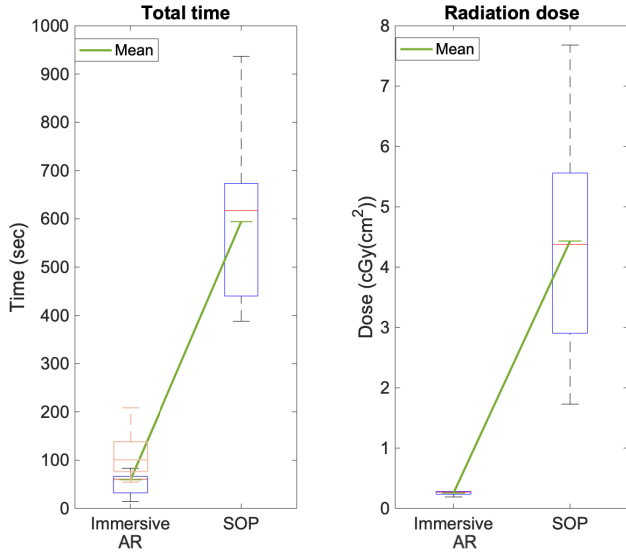


Fig. 13. The plots present the execution time and total radiation dose during K-wire insertion using the AR supported approach and SOP. On the leftmost plot, the blue boxplot is the execution time with AR, whereas the orange boxplot is the total time including the planning phase. The green lines show the mean values for each of the groups.

and the abduction and anteversion angles of the acetabular implant, based on standard guidelines. Table III shows the results from the baseline standard operating procedure (SOP) using conventional fluoroscopic guidance.

Table IV provides a comparison of the mean and standard deviation (SD) values of the K-wire insertion errors using our immersive AR system with a previous non-immersive AR system [22] as well as the SOP. Combining the planning and execution times, the AR procedure took on average 111.25 sec versus the 594.3 sec during SOP. Fig. 13 depicts this comparison. On average, the surgeons used 2 fluoroscopic shots with a combined dose of $0.255 \text{ cGy}(\text{cm}^2)$ per user and committed an insertion error of 4.76 mm. Given the eight samples, the population mean for the drilling error, based on the 95% confidence intervals, is between $[2.60 - 6.92] \text{ mm}$.

Correspondingly, we present the outcome for the acetabular cup placement procedure with the immersive AR system in Table V, comparing it to a previous non-immersive AR application [44] and SOP. Using SOP, it took surgeons on average 235 sec to place the cup and under AR a combined time of 122.38 sec was achieved. For the AR setup, we acquired 8 X-ray images with an average dose of $1.25 \text{ cGy}(\text{cm}^2)$ per surgeon, whereas 14 fluoroscopic images with a dose of $1.96 \text{ cGy}(\text{cm}^2)$ were acquired during SOP. With the AR system, the average errors were 1.57° and 1.46° for the abduction and anteversion angles, respectively. Based on the 95% confidence interval, the mean error for abduction and anteversion are between $[0.72^\circ - 2.44^\circ]$ and $[0.72^\circ - 2.20^\circ]$, respectively. Under SOP the respective angles were 4.76° and 4.78° . Figs. 14 and 15 present the outcome with respect to time, radiation dose, and individual rotational measures for the acetabular cup placement experiments using AR and SOP. The immersive AR results show an SD of respectively 89.88 sec and 32.5 sec for planning and execution time, 0 for the

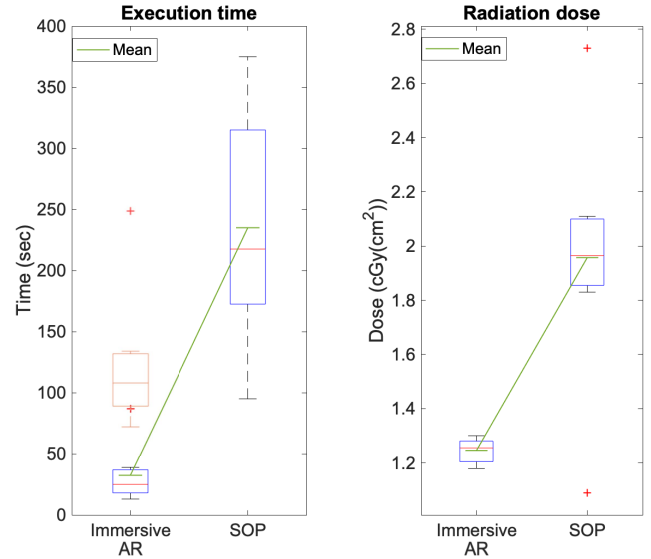


Fig. 14. Comparison of time and total radiation dose during cup placement with AR and SOP approaches. The orange boxplot represents the total time including the planning time. The red (+) denote outliers, where in the leftmost plot the top sign belongs to the orange boxplot, and the bottom (+) to the blue plot.

number of X-ray images, 1.25 for the dose, 1.24 for the abduction error and 1.07 in the anteversion error. Comparable to Tables I and II, Table III displays the individual participants performance during each of the two SOP. To statistically evaluate our findings, a two sample t-test was performed and the results are reported in Table VI. The threshold for statistical significance is considered as $p = 0.05$. We tested the results of our immersive AR system against the results from the NI-AR system and SOP for both the K-wire insertion procedure and the acetabular cup placement. We did not test the number of acquired X-rays, since in these experiments they turned out to be a constant value for each procedure.

IV. DISCUSSION

We evaluated our spatially-aware AR system in two clinically relevant procedures, *i)* the placement of K-wires through tubular structures for fracture repair tasks, and *ii)* placement of acetabular components into the hip socket for total hip arthroplasty. We selected these two high volume procedures among the many other applications which can be enabled by our interactive AR system, as they each represent a class of common orientational alignment and localization tasks that are prevalent across different fields of image-guided surgery.

For the K-wire insertion procedure, the immersive AR system performed significantly faster than the conventional SOP, yielding less than a fifth of the time (Fig. 13). Table IV demonstrates a detailed comparison of our system, not only with the SOP as an established baseline, but also with a previously presented non-immersive mixed reality method based on RGBD sensing and intra-operative CBCT imaging [22].

With the AR system every surgeon used exactly 2 X-ray images, which were the 2 images required for procedure planning. Despite explicitly allowing them to take as many

TABLE VI

RESULTS FROM TWO SAMPLE T-TESTS OF OUR AR METHOD COMPARED TO THE RESPECTIVE NI-AR METHOD AND THE SOP. THE UPPER HALF OF THE TABLE SHOWS RESULTS OF THE STATISTICAL EVALUATION OF K-WIRE INSERTION WITH THE AR AND NI-AR VERSION AS WELL AS AR AND SOP. THE LOWER HALF CONTAINS THE CORRESPONDING VALUES FOR THE ACETABULAR CUP PLACEMENT

	P-value	Time (sec)	Dose (cGY(cm ²))	Error (mm)	Abd. (°)	Ant. (°)
K-wire	AR / NI-AR [22]	$3.27 \cdot 10^{-5}$	$1.25 \cdot 10^{-11}$	0.81	-	-
	AR / SOP	$0.95 \cdot 10^{-5}$	$4.91 \cdot 10^{-5}$	0.93	-	-
THA	AR / NI-AR [44]	$1.64 \cdot 10^{-6}$	0.34	- -	0.49	0.22
	AR / SOP	$0.12 \cdot 10^{-1}$	$5.39 \cdot 10^{-4}$	- -	$0.31 \cdot 10^{-2}$	$0.14 \cdot 10^{-1}$

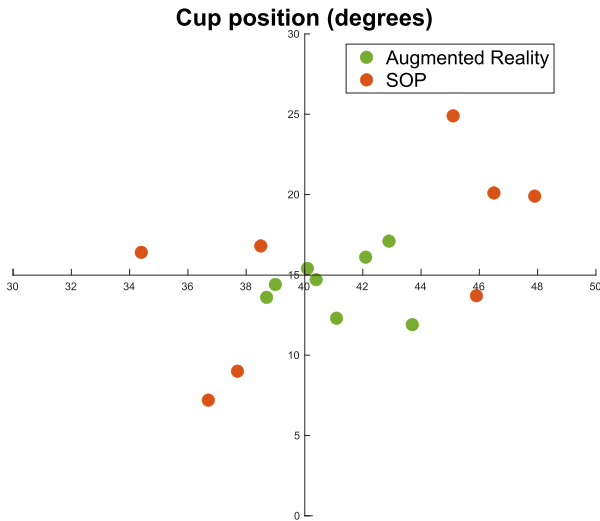


Fig. 15. Anteversion and abduction angles are shown after acetabular cup placement using AR support and SOP. The horizontal axis represents the abduction angle, and the vertical axis represents the anteversion—the center of the plot corresponds to the desired angles of 40° and 15°. The farther the data points from the center signify higher errors committed by the user. The AR method resulted in a stronger cluster near the center, while SOP yielded higher errors and more outliers.

radiographs as they desire, no one of the surgeons requested additional X-ray images. As mentioned above, during SOP, surgeons inserted the K-wires with an average number of 40.86 fluoroscopic images and with an average dose of 4.43 cGY(cm²), compared to the (statistically) significantly lower dose of 0.255 cGY(cm²), which was emitted during the AR procedures. The RGBD-CBCT system in [22], [23] yielded on average 2.14 X-rays, although it required a pre-procedural CBCT scan of the phantom, inducing the higher radiation dose of 1.6 cGY(cm²).

Finally, evaluating the outcome of the procedure with regard to the drilling error, AR (4.76 mm) outperforms RGBD-CBCT (5.13 mm), both being marginally worse than SOP (4.61 mm). The differences in accuracy were not statistically significant. Considering that we only instructed the surgeons to drill through the tube and not precisely through the center of the tube, we regard these difference as negligible. It is important to note that, our AR system performed similar to the conventional X-ray method in terms of accuracy, while reducing time by a factor of 5, number of fluoroscopic acquisitions by a factor of 20, and the radiation dose by a factor of 17.

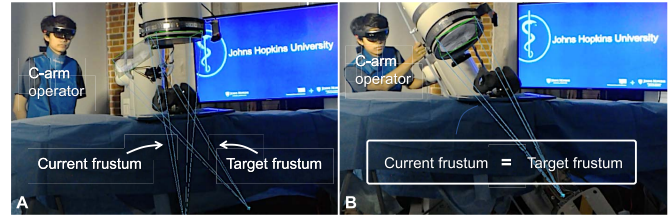


Fig. 16. Visualization of a target frustum (A) allows the C-arm operator to align the current C-arm frustum with the surgeon’s desired perspective (B) and eliminate the waste of time and radiation during fluoro hunting. This concept is an example of the capabilities of interactive frustums on moving information between different stake holders in the OR, *i.e.* surgeon, patient, X-ray technician, staff, etc.

We observed that, in addition to the planning information, the surgeons took multiple other considerations into account while deciding on the insertion path. One of which was direct visualization of the X-ray images that they acquired for planning. Observing the C-arm pose with respect to the patient, and the visualization of the images within their viewing frustum, assisted them in better localizing the target structure. In our setup, tactile feedback did not play a significant role; however, in reducing real fractures, haptic feedback further assists the surgeon in identifying whether the wire is inserted in specific anatomy or not.

The SLAM-based error for HMD is dominantly along the principal axis of the viewer, *i.e.*, direction pointing away from the user to the scene. When the users are presented with a trajectory to follow, we observed that they naturally place their heads in the direction of the target trajectory for optimal visualization and ease of alignment (Fig. 3). This configuration is advantageous since the maximum tracking uncertainty is in the direction of the drilling trajectory, which is the direction that is not relevant to the drilling task. This is because the amount of penetration of the drill is easily identified from X-ray images. This uncertainty behaviour in AR has been previously investigated by Hoff *et al.* [46].

A similar trend to the K-wire experiment is observed with the measurements for the placement of the acetabular cup, demonstrating the effectiveness of our AR system, we compare it against SOP and a NI-AR system as presented in [44]. As shown in Fig. 14, the execution time is considerably lower using AR; even when combining planning and execution time, it took the surgeons 122.38 sec, which is nearly half of the 235 sec that they needed under SOP and comparable to

the 110.6sec with NI-AR. Both differences are statistically significant. Furthermore, the number of fluoroscopic images were reduced; every surgeon used exactly 8 images, which are again merely the images required for planning. This resulted in an average dose of $1.25\text{cGY}(\text{cm}^2)$, which is significantly lower than with SOP, where the surgeons used an average of 13.75 radiographs with an average dose of $1.96\text{cGY}(\text{cm}^2)$, and lower than with NI-AR where one pre-procedure CBCT lead to a dose of $1.83\text{cGY}(\text{cm}^2)$. The objective of this procedure was to achieve abduction and anteversion angles of 40° and 15° , respectively, which lie in the clinical safe-zone [45]. The respective errors are shown in Table V and Fig. 15. The outcome distinctly displays a more accurate cup placement using the spatially-aware immersive AR system (1.57° & 1.46°) compared to the SOP (4.76° & 4.78°), compared to the NI-AR system (1.78° & 1.43°) the abduction error is slightly less, whereas the anteversion error is marginally higher (0.03°). The differences in the abduction and anteversion errors between the immersive AR system and SOP are both statistically significant.

Concerning the planning and execution time, some participants such as **P2**, **P6**, and **P8**, who had previous experience with AR HMDs and were familiar with our system performed faster than other users, especially compared to **P1** and **P7**, who had no prior exposure to HMDs. In order to reduce the disparity in acquaintance with the technology, each participant completed a short training in which they were familiarized with the headset and interaction techniques used in our software. Despite the same training session, the differences mentioned above still seem related to experience with AR. We expect the performances to level after users gain more experience with the system. These outcomes are not surprising as any new technology requires time and experience to exploit its full potential.

In almost all orthopedic procedures, we rarely use live fluoroscopy and rely almost exclusively on still images. Therefore, the frame-rate for these procedures appeared entirely adequate. The resolution also seemed satisfactory for the experiments that we performed. No users complained about not being able to “see the planning annotations”. However, if the system wanted to be adapted to other fields, such as interventional neuroradiology — the suitability of resolution, contrast, brightness, frame-rate, etc. may need to be further investigated.

For both procedures the deployment of our AR system lead to a comparable or higher accuracy, fewer X-ray images with a consequently lower radiation dose. For the total time, it has to be noted that our planning time does not include the recording of the X-ray images that were necessary to plan the procedure. This step however can be fully automated, resulting in an immediate availability of the fluoroscopic images.

V. CONCLUSION AND FUTURE WORK

We presented the embodiment of a novel interaction concept based on spatiotemporal-aware AR. In our work, we aimed to provide meaningful registration and visualization without the need for tracking patients or tools with outside-in navigation

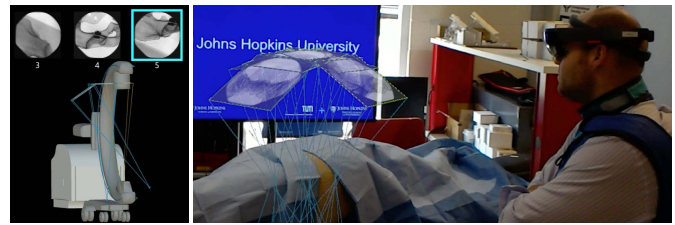


Fig. 17. Spatial and temporal information from the surgery can be recorded and reviewed after surgery. On the left side an interface is shown allowing the surgeon to select images, which he can then observe geometrically accurate in space, as shown on the right side.

hardware. Instead, we brought intuition to visualization by connecting the viewing frustums of the scanner with the surgical team. For the two orthopedic use cases presented in this manuscript, our immersive AR system demonstrated improvements in time, number of X-ray acquisitions, radiation dose, and outcome during cup placement.

The most significant source of error is from the localization of the AR HMD using SLAM. This error is present both during calibration and application. If additional sensors such as IMUs and depth cameras would constraint the tracking algorithm in the future, we can expect improved tracking and smaller drifts.

As computer vision methodologies enter the medical and AR domains, the accuracy and robustness of medical AR systems will also increase. This work presents a general end-to-end workflow for integration of AR into orthopedic surgery, which will also observe dramatic improvement as such progress occurs.

The spatiotemporal awareness inherent in AR overhauls the ill-posed communication between the surgeon, staff, and information; e.g. Fig. 16 shows the potential role of flying frustums and AR in effectively communicating desired X-ray views to the technician, eliminating unfavorable views and reducing the staff burnout.

A significant advantage of our system is that it is based on 2D C-arm fluoroscopy — and therefore does not diverge from the standard workflow. The standard C-arm is always present in the OR to take confirmatory images. Consequently, the surgeons are not required to solely rely on the AR system during a procedure. Rather, its value lies in guiding the surgeon to a narrow area of interest and more importantly, to align them with the right trajectory. The surgeon can then seamlessly switch between standard fluoroscopic images and the AR view to guarantee accurate drilling. We believe this is useful, translatable, safe, and novel.

An application of AR outside the OR is “surgical replay”, where the residents can review the surgery, accompanied with its temporal and spatial information including all the X-ray acquisitions and optical point-clouds from the patient site. This enables the medical trainees to identify distinct actions that were taken by the experienced surgeon based upon each image. Access to such 3D post-operative analysis has the potential to dramatically improve the quality of surgical education.

In a training environment, such as in most academic centers with residents and fellows — this ability for all users to wear the HMD and view the same annotations and AR guidance

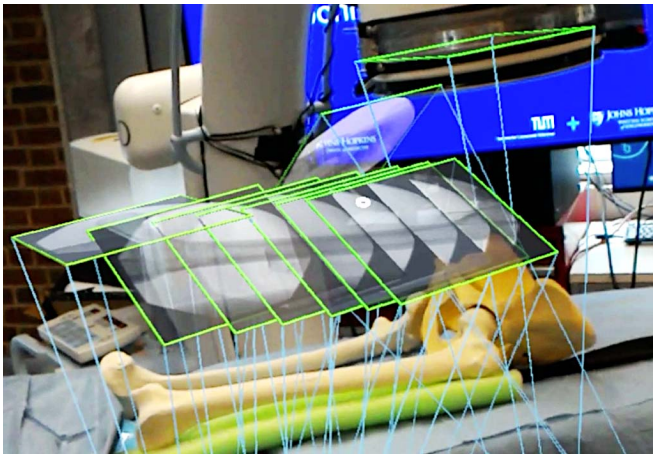


Fig. 18. The interlocking of multiple X-ray frustums enables visualization of large anatomical structures. In this figure, multiple images are acquired on a co-linear trajectory and are locked to each other to form a quasi-panoramic view of the bone.

would be helpful. The more experienced user could ensure that the anatomy had been annotated appropriately and that the AR guidance trajectories are at the correct location.

Direct visualization of X-ray images within their corresponding viewing frustums delivers intuition that effectively unites the content of the 2D image with the 3D imaged anatomy. In this setting, images from various perspectives can be grouped within their frustums to form multi- or extended-view representations of the anatomy. The interlocked frustums shown in Fig. 18 are examples for such visualization concept, that can particularly benefit interventions where leg-length discrepancies or malrotations in tibio- and lateral/distal-femoral angles are major concerns.

In this work, we demonstrated that two elements share the AR experience; the first one was the augmented surgeon, and the second one was the imaging observer which was the C-arm scanner that used a calibrated HMD attached to it. With the same design, any number of other users can connect to the same AR session, collaborate, and share an experience. Though assessment of each of these concepts presented in Figs. 16, 17, and 18 requires an additional approved study, we believe their introduction to the community paves the way in opening new paths for research in this area and expedites the translation of AR-based solutions into future ORs.

Despite that our solution delivers spatial awareness, it should not be regarded as a surgical navigation system. This is because marker-less tracking, currently, cannot deliver the level of accuracy achieved by marker-based surgical navigation or robotic systems. It is important to note that, as most AR systems display the data anchored to the surrounding environment, tracking with respect to the environment has become an inherent component in modern AR systems. In an optical see-through HMD, the integrated camera sensors are used to run SLAM in the background, construct a dynamic map from the environment, and localize the HMD within that environment. Despite that this marker-less mapping and localization cannot deliver the same level of accuracy as the external navigation systems with explicit tracking nature, it delivers spatial aware-

ness for the visualization of information. Hence, we regard our solution as a spatially-aware advanced visualization platform that enhances the interaction across the surgical ecosystem and promotes effective collaboration and team approach.

The widespread adoption of human-centered AR in interventional routines requires careful considerations regarding surgeons' experience. The interaction with the virtual content using intuitive hand gestures, and resolving the perceptual ambiguities between real and virtual that occur due to vergence-accommodation conflict, can greatly contribute to the acceptance of AR technologies. Furthermore, development of artificial intelligence strategies can *i)* create semantic understanding from the surgical environment and augment surgeon's intelligence, and *ii)* enhance the spatial mapping and co-localization, thus improving the stability of marker-less AR systems. Lastly, future integration of eye tracking systems into HMDs can circumvent the internal eye-to-display calibration, adjust the rendering based on each user, and replace unnecessary gestures with gaze-based interaction.

Regarding sterilization, since the HMD on the user's head does not enter the operating field, it could be viewed in a similar way to surgical loupes, surgical hoods, headlamps — all of which are used daily and do not need to be sterilized.

Our solution that promotes the team approach, also demands the entire crew to get trained appropriately and be able to interact with the system comfortably. Therefore, time, money, and effort need to be spent to prepare the surgical team for such a digital transformation. From a hardware standpoint, other than HMDs, no additional hardware is required in the OR. Therefore, the device cost is lower than complex navigation-based or robotic solutions.

As shown in this paper, the introduction of new technology requires a user-centric design for its full integration into the clinical workflow. The ultimate goal may be the discovery of new surgical workflows enabled by the introduction of novel technology. This goal could only be achieved with a close and extended partnership between surgeons and technical experts, as well as the full integration of the advanced technology in the broad spectrum of surgical teaching, training, planning, workflow, and documentation.

REFERENCES

- [1] J. H. Siewerdsen *et al.*, "Volume CT with a flat-panel detector on a mobile, isocentric C-arm: Pre-clinical investigation in guidance of minimally invasive surgery," *Med. Phys.*, vol. 32, no. 1, pp. 241–254, Jan. 2005.
- [2] J. S. Hott *et al.*, "Intraoperative Iso-C C-Arm navigation in craniocervical surgery: The first 60 cases," *Neurosurgery*, vol. 54, no. 5, pp. 1131–1137, May 2004.
- [3] D. L. Miller *et al.*, "Occupational radiation protection in interventional radiology: A joint guideline of the cardiovascular and interventional radiology society of europe and the society of interventional radiology," *CardioVascular Interventional Radiol.*, vol. 33, no. 2, pp. 230–239, Apr. 2010.
- [4] N. Theocharopoulos, K. Perisinakis, J. Damilakis, G. Papadokostakis, A. Hadjipavlou, and N. Gourtsoyiannis, "Occupational exposure from common fluoroscopic projections used in orthopaedic surgery," *J. Bone Joint Surg.-Amer. Volume*, vol. 85, no. 9, pp. 1698–1703, Sep. 2003.
- [5] C. W. Kim, Y.-P. Lee, W. Taylor, A. Oygur, and W. K. Kim, "Use of navigation-assisted fluoroscopy to decrease radiation exposure during minimally invasive spine surgery," *Spine J.*, vol. 8, no. 4, pp. 584–590, Jul. 2008.

- [6] U. K. Kakarla, A. S. Little, S. W. Chang, V. K. H. Sonntag, and N. Theodore, "Placement of percutaneous thoracic pedicle screws using NeuroNavigation," *World Neurosurg.*, vol. 74, no. 6, pp. 606–610, Dec. 2010.
- [7] N. R. Crawford, N. Theodore, and M. A. Foster, "Surgical robot platform," U.S. Patent 9782229, Oct. 10, 2017.
- [8] T. Yi, V. Ramchandran, J. H. Siewersden, and A. Uneri, "Robotic drill guide positioning using known-component 3D–2D image registration," *J. Med. Imag.*, vol. 5, no. 2, Feb. 2018, Art. no. 021212.
- [9] L. Jaskowicz and E. J. Hazan, "Computer aided orthopaedic surgery: Incremental shift or paradigm change?" *Med. Image Anal.*, vol. 33, pp. 84–90, Oct. 2016.
- [10] R. B. Grupp *et al.*, "Pose estimation of periacetabular osteotomy fragments with intraoperative X-ray navigation," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 441–452, Feb. 2020.
- [11] J. Goerres *et al.*, "Planning, guidance, and quality assurance of pelvic screw placement using deformable image registration," *Phys. Med. Biol.*, vol. 62, no. 23, p. 9018, 2017.
- [12] M. Synowitz and J. Kiwit, "Surgeon's radiation exposure during percutaneous vertebroplasty," *J. Neurosurg., Spine*, vol. 4, no. 2, pp. 106–109, Feb. 2006.
- [13] A. J. Starr, A. L. Jones, C. M. Reinert, and D. S. Borer, "Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the Acetabulum," *Injury*, vol. 32, pp. 45–50, May 2001.
- [14] N. Navab, J. Traub, T. Sielhorst, M. Feuerstein, and C. Bichlmeier, "Action- and workflow-driven augmented reality for computer-aided medical procedures," *IEEE Comput. Graph. Appl.*, vol. 27, no. 5, pp. 10–14, Sep. 2007.
- [15] K. Aagaard, B. S. Laursen, B. S. Rasmussen, and E. E. Sørensen, "Interaction between nurse anesthetists and patients in a highly technological environment," *J. PeriAnesthesia Nursing*, vol. 32, no. 5, pp. 453–463, Oct. 2017.
- [16] C. Laverdière *et al.*, "Augmented reality in orthopaedics: A systematic review and a window on future possibilities," *Bone Joint J.*, vol. 101-B, no. 12, pp. 1479–1488, Dec. 2019.
- [17] B. Hatscher, M. Luz, L. E. Nacke, N. Elkmann, V. Müller, and C. Hansen, "GazeTap: Towards hands-free interaction in the operating room," in *Proc. 19th ACM Int. Conf. Multimodal Interact. ICMI*, 2017, pp. 243–251.
- [18] A. Mewes, B. Hensen, F. Wacker, and C. Hansen, "Touchless interaction with software in interventional radiology and surgery: A systematic literature review," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 291–305, Feb. 2017.
- [19] Y. Sato *et al.*, "Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 681–693, Oct. 1998.
- [20] N. Navab, A. Bani-Kashemi, and M. Mitschke, "Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications," in *Proc. 2nd IEEE ACM Int. Workshop Augmented Reality (IWAR)*, Oct. 1999, pp. 134–141.
- [21] N. Navab, S.-M. Heining, and J. Traub, "Camera augmented mobile C-Arm (CAMC): Calibration, accuracy study, and clinical applications," *IEEE Trans. Med. Imag.*, vol. 29, no. 7, pp. 1412–1423, Jul. 2010.
- [22] M. Fischer *et al.*, "Preclinical usability study of multiple augmented reality concepts for K-wire placement," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 1007–1014, Jun. 2016.
- [23] J. Fotouhi *et al.*, "Interventional 3D augmented reality for orthopedic and trauma surgery," in *Proc. 16th Annu. Meeting Int. Soc. Comput. Assist. Orthopedic Surg. (CAOS)*, 2016, pp. 1–6.
- [24] J. Fotouhi *et al.*, "Pose-aware C-arm for automatic re-initialization of interventional 2D/3D image registration," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 7, pp. 1221–1230, Jul. 2017.
- [25] J. Fotouhi *et al.*, "Plan in 2-d, execute in 3-D: An augmented reality solution for cup placement in total hip arthroplasty," *J. Med. Imag.*, vol. 5, no. 2, 2018, Art. no. 021205.
- [26] F. Müller, S. Roner, F. Liebmann, J. M. Spirig, P. Föhnstahl, and M. Farshad, "Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging," *Spine J.*, vol. 20, no. 4, pp. 621–628, Apr. 2020.
- [27] C. A. Agten, C. Dennler, A. B. Roskopf, L. Jaberg, C. W. A. Pfirrmann, and M. Farshad, "Augmented Reality–Guided lumbar facet joint injections," *Investigative Radiol.*, vol. 53, no. 8, pp. 495–498, Aug. 2018.
- [28] J. T. Gibby, S. A. Swenson, S. Cvetko, R. Rao, and R. Javan, "Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 3, pp. 525–535, Mar. 2019.
- [29] B. H. van Duren, K. Sugand, R. Wescott, R. Carrington, and A. Hart, "Augmented reality fluoroscopy simulation of the guide-wire insertion in DHS surgery: A proof of concept study," *Med. Eng. Phys.*, vol. 55, pp. 52–59, May 2018.
- [30] H. S. Cho *et al.*, "Can augmented reality be helpful in pelvic bone cancer surgery? An *in vitro* study," *Clin. Orthopaedics Rel. Res.*, vol. 476, no. 9, pp. 1719–1725, Sep. 2018.
- [31] H. Ogawa, S. Hasegawa, S. Tsukada, and M. Matsubara, "A pilot study of augmented reality technology applied to the acetabular cup placement during total hip arthroplasty," *J. Arthroplasty*, vol. 33, no. 6, pp. 1833–1837, Jun. 2018.
- [32] H. Brun *et al.*, "Mixed reality holograms for heart surgery planning: First user experience in congenital heart disease," *Eur. Heart J. Cardiovascular Imag.*, vol. 20, no. 8, pp. 883–888, Aug. 2019.
- [33] P. E. Pelargos *et al.*, "Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery," *J. Clin. Neurosci.*, vol. 35, pp. 1–4, Jan. 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," *Plastic Reconstructive Surg.*, vol. 136, no. 2, pp. 328–330, Aug. 2015.
- [35] R. Moreta-Martinez, D. García-Mato, M. García-Sevilla, R. Pérez-Mañanes, J. Calvo-Haro, and J. Pascau, "Augmented reality in computer-assisted interventions based on patient-specific 3D printed reference," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 162–166, Oct. 2018.
- [36] J. W. Meulstee *et al.*, "Toward holographic-guided surgery," *Surgical Innov.*, vol. 26, no. 1, pp. 86–94, Feb. 2019.
- [37] L. Ma *et al.*, "Three-dimensional augmented reality surgical navigation with hybrid optical and electromagnetic tracking for distal intramedullary nail interlocking," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 4, Aug. 2018, Art. no. e1909.
- [38] S. Andress *et al.*, "On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial," *J. Med. Imag.*, vol. 5, no. 2, Jan. 2018, Art. no. 021209.
- [39] J. Hajek *et al.*, "Closing the calibration loop: An inside-out-tracking paradigm for augmented reality in orthopedic surgery," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2018, pp. 299–306.
- [40] J. Fotouhi *et al.*, "Co-localized augmented human and X-ray observers in collaborative surgical ecosystem," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 9, pp. 1553–1563, Sep. 2019.
- [41] J. Fotouhi *et al.*, "Interactive flying frustums (IFFs): Spatially aware surgical data visualization," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 6, pp. 913–922, 2019.
- [42] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [43] K. T. Foley, D. A. Simon, and Y. R. Rampersaud, "Virtual fluoroscopy: Computer-assisted fluoroscopic navigation," *Spine*, vol. 26, no. 4, pp. 347–351, Feb. 2001.
- [44] C. Alexander, A. E. Loeb, J. Fotouhi, N. Navab, M. Armand, and H. S. Khanuja, "Augmented reality for acetabular component placement in direct anterior total hip arthroplasty," *J. Arthroplasty*, vol. 35, no. 6, pp. 1636–1641, Jun. 2020.
- [45] G. E. Lewinnek, J. L. Lewis, R. Tarr, C. L. Compere, and J. R. Zimmerman, "Dislocations after total hip-replacement arthroplasties," *J. Bone Joint Surg. Amer.*, vol. 60, no. 2, pp. 217–220, 1978.
- [46] W. Hoff and T. Vincent, "Analysis of head pose accuracy in augmented reality," *IEEE Trans. Vis. Comput. Graphics*, vol. 6, no. 4, pp. 319–334, Oct. 2000.