Interactive initialization of 2D/3D rigid registration

Ren Hui Gong and Özgür Güler
The Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington, DC 20010

Mustafa Kürklüoglu
Department of Cardiac Surgery, Children's National Medical Center, Washington, DC 20010

John Lovejoy
Department of Orthopaedic Surgery and Sports Medicine, Children's National Medical Center, Washington, DC 20010

Ziv Yaniv[a)]
The Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington, DC 20010 and Departments of Pediatrics and Radiology, George Washington University, Washington, DC 20037

(Received 14 July 2013; revised 9 October 2013; accepted for publication 31 October 2013; published 19 November 2013)

Purpose: Registration is one of the key technical components in an image-guided navigation system. A large number of 2D/3D registration algorithms have been previously proposed, but have not been able to transition into clinical practice. The authors identify the primary reason for the lack of adoption with the prerequisite for a sufficiently accurate initial transformation, mean target registration error of about 10 mm or less. In this paper, the authors present two interactive initialization approaches that provide the desired accuracy for x-ray/MR and x-ray/CT registration in the operating room setting.
Methods: The authors have developed two interactive registration methods based on visual alignment of a preoperative image, MR, or CT to intraoperative x-rays. In the first approach, the operator uses a gesture based interface to align a volume rendering of the preoperative image to multiple x-rays. The second approach uses a tracked tool available as part of a navigation system. Preoperatively, a virtual replica of the tool is positioned next to the anatomical structures visible in the volumetric data. Intraoperatively, the physical tool is positioned in a similar manner and subsequently used to align a volume rendering to the x-ray images using an augmented reality (AR) approach. Both methods were assessed using three publicly available reference data sets for 2D/3D registration evaluation.
Results: In the authors' experiments, the authors show that for x-ray/MR registration, the gesture based method resulted in a mean target registration error (mTRE) of 9.3 ± 5.0 mm with an average interaction time of 146.3 ± 73.0 s, and the AR-based method had mTREs of 7.2 ± 3.2 mm with interaction times of 44 ± 32 s. For x-ray/CT registration, the gesture based method resulted in a mTRE of 7.4 ± 5.0 mm with an average interaction time of 132.1 ± 66.4 s, and the AR-based method had mTREs of 8.3 ± 5.0 mm with interaction times of 58 ± 52 s.
Conclusions: Based on the authors' evaluation, the authors conclude that the registration approaches are sufficiently accurate for initializing 2D/3D registration in the OR setting, both when a tracking system is not in use (gesture based approach), and when a tracking system is already in use (AR based approach). © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4830428]

Key words: image registration, image guidance, gesture based interaction, augmented reality

1. INTRODUCTION

Registration of preoperative images to the interventional setting is one of the key technical components of image-guided navigation systems.[1] The most commonly utilized registration approaches rely on the use of fiducials or anatomical landmarks and surfaces. These structures are localized in the preoperative image either manually or automatically. Intraoperatively, they are digitized using a tracked pointer tool or, possibly, a laser range scanner.[2] The preoperative 3D data are then registered to the 3D data acquired intraoperatively using various 3D/3D point-based rigid registration algorithms.[3]

An alternative approach is to perform 2D/3D anatomy-based rigid registration, aligning the preoperative volume, MR, or CT to the intraoperative setting using a few x-ray images. This subject has been studied extensively, resulting in a large number of published algorithms. A broad overview of the various approaches to solving this challenge is given in a recent survey.[4] While the differences are many, all of these algorithms have one thing in common, they are iterative and require initialization.

In practice, the majority of these algorithms have not been able to transition from bench to bedside. The only domain where 2D/3D registration has become part of standard

clinical care is radiation therapy. The distinguishing characteristic of this domain is that a good initial estimate of the transformation is obtained by accurate patient positioning using other means.

Various initialization approaches have been used in the context of 2D/3D registration.[3,4] The three most common methods are as follows:

1. Manual initialization — interactively manipulate the pose of the 3D image using the keyboard and mouse. In combination with the known x-ray camera parameters, a 2D image is generated from the volumetric image. The user actively manipulates the transformation parameters such that the computer generated image and the x-ray image become visually similar.

   It should be noted that there are clinical procedures where such an approach can serve as the final registration result. Two recent examples include x-ray/MR registration for congenital heart disease catheterization[5] and x-ray/CT registration for radiation therapy treatment of head and neck cancer.[6]
2. Coarse paired point registration — using an analytic, paired-point registration algorithm with coarsely localized points in the intraoperative setting. These can be either skin adhesive fiducials or anatomical landmarks. Points can be localized intraoperatively either by using a tracked pointer tool or by using several x-ray images and estimating point locations as the intersection of back-projected rays. A recent example describing automatic identification of a sparse set of corresponding anatomical landmarks, vessel bifurcations, and spine centerline to perform initialization was described in Ref. 7.

   When using fiducials, it is sometimes possible to accurately localize them both in the x-ray and the 3D imaging modality, and thus obtain the final registration result. An example illustrating this approach for x-ray/MR registration is described in Ref. 8, and Ref. 9 describes this approach in the context of x-ray/CT registration.
3. Clinical setup — using the known geometry of the intraoperative imaging system to bound the transformation parameters. A rough initialization can be obtained by using the intersection point of all principle rays to position the preoperative image.[10] Procedure specific knowledge can also be incorporated to estimate orientation. That is, assuming a specific patient setup (e.g., supine), and associating x-ray images with specific camera orientations (e.g., anterior-posterior).[11]

Additional initialization approaches include:

1. Brute force[12] — using an estimate of the transformation parameter values based on clinical setup, sample the parameter space in that region using a coarse grid. For all parameter values generate corresponding digitally reconstructed radiographs and compute the similarity measure's value. The parameters corresponding to the best similarity value are used as the initial estimate.
2. 3D/3D intensity based registration[13] — obtain an initial registration by registering intraoperative Cone Beam CT (CBCT) to the preoperative CT. This approach is also readily applicable to initialization of x-ray/MR registration, via CBCT/MR registration.
3. Fourier slice theorem[14] — specifically address the task of initialization in the context of x-ray/CT registration. In this work, the Fourier slice theorem facilitates estimation of the orientation, and phase correlation is used to estimate translation. This method was able to improve the registration success rate of an intensity based registration algorithm from 28.8% to 68.6%.
4. Virtual fiducial marker[15] — using a bootstrap form of approach to initialize x-ray/CT registration, equally applicable to x-ray MR. This method is based on the use of a virtual fiducial marker. That is, an easily identifiable set of fiducials, graduations of radiopaque ruler, are placed next to the patient intraoperatively. In an initial setup step, two wide field of view x-rays are acquired and x-ray/CT registration is performed using the paired-point approach with corresponding anatomical landmarks localized in the two x-rays and the CT. The ruler coordinates are then mapped to the CT coordinate system, creating a virtual fiducial. From here on, reinitialization is obtained in part using knowledge about the clinical setup, orientation, and one component of translation, and use of a single virtual fiducial point and triangulation to resolve the two additional translational components. The method was successfully evaluated using data from 31 aortic aneurysm repair interventions.

In this work, we present two interactive approaches for initialization of 2D/3D registration. The first is based on the use of gestures to visually align a volume rendering of the anatomical structures with x-ray images. The second is based on the use of a tracked tool to visually align the volume with the x-ray images using an augmented reality (AR) approach. The former is applicable in procedures where a tracking system is not part of the current setup, while the later is relevant in procedures where a tracking system is already available.

2. MATERIALS AND METHODS

2.A. Interactive initialization

2.A.1. Gesture-based initialization

We have developed a graphical user interface which allows the operator to interactively position a 3D volume in space, with the intent of aligning it to an x-ray image. Interaction with our program is based on a predefined set of gestures, enabling touchless interaction, which in turn allows the surgeon to control the process while maintaining sterility. Figure 1 shows the system setup with the Microsoft Kinect.

The user starts by loading a volumetric data set, MR, or CT, and a set of x-ray images. Initialization is then performed in two steps: automatic positioning limited to translation, and interactive pose determination.

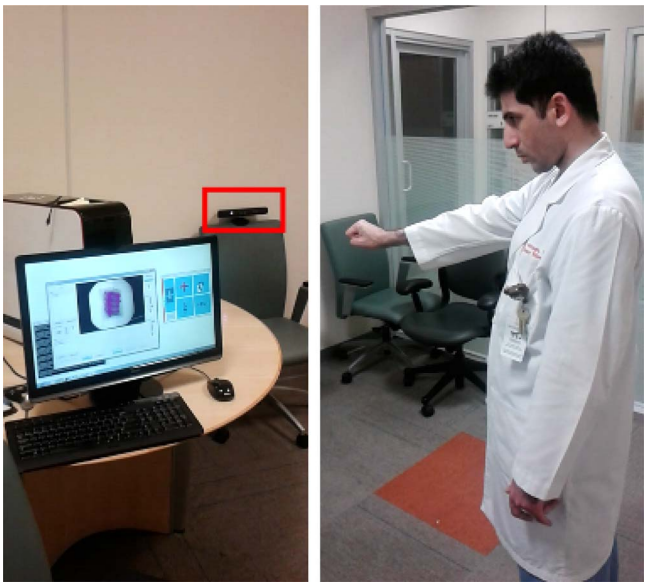

FIG. 1. Setup of the gesture based approach. (left) Kinect device and computer screen displaying the graphical user interface and interaction panel. (right) user performing a grip gesture.

2.A.1.a. Automatic positioning. Initial positioning of the 3D data is performed by aligning the geometrical center of the volumetric data set with the intersection point of the principal rays from all x-ray cameras. In practice, the rays usually do not intersect at a single point, thus an averaged intersection point is calculated. To obtain this point, for every pair of x-ray principal rays we find the shortest line segment between the two rays, and get the two intersecting points. This will produce a cloud of points, whose center is used as the intersection point. This positioning only accounts for translation, and has minimal requirements with regard to accuracy. That is, we only require that the volume rendering overlap with all of the x-ray images such that the user can interact with it in the following step.

2.A.1.b. Interactive pose determination. In this step, the operator refines the gross estimate obtained by the automatic positioning. They interactively modify the position and orientation of the volume rendering with the goal of increasing the visual overlap with the set of x-ray images. This is done in an iterative manner, with the operator selecting an x-ray image, manipulating the volume's pose, and then switching to another x-ray image. The process continues until the operator is satisfied by the overlaps obtained for all of the images.

Previously we had implemented this approach relying solely on the graphical user interface shown on the left of Fig. 2.[10] Interaction was done using four different gestures, and required the use of both hands. In addition, the user did not receive any feedback with regard to the detected gesture and current control mode. This incurred a high mental burden and required a learning phase prior to interacting with the system. Based on these observations, we have modified our interaction approach as follows. We have added an interaction panel, shown in Fig. 2, and switched to using intuitive one handed gestures, grip (closed fist) and release (open hand). In our current work, we only used the right hand, since all the participants were right handed. Adjusting to left handed interaction is straight forward.

Our interaction panel consists of information panes, and control panes. The information panes indicate the user's hand location within the Kinect's working volume. Gesture detection is performed only if the operators right hand is detected inside the active work volume (at a distance of 1.8 m−2.8 m from the Kinect). All control panes are activated by the grip gesture and deactivated by the release gesture. The cursor

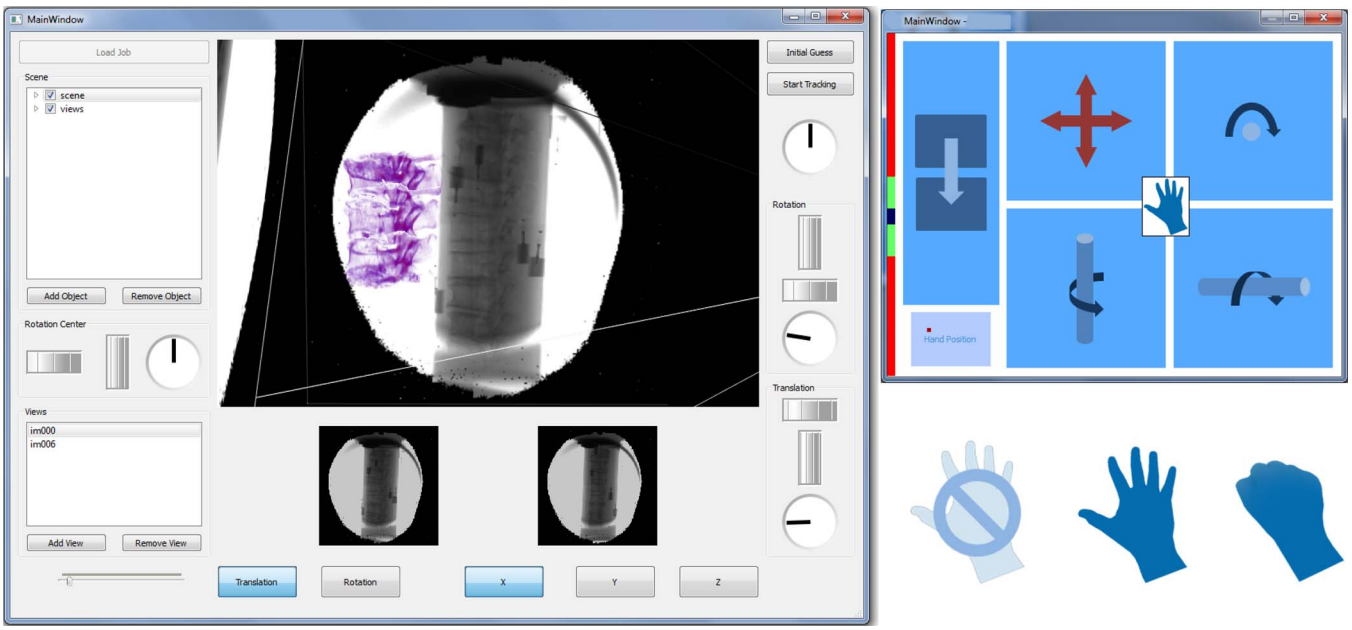

FIG. 2. (Left) Graphical user interface for gesture-based 2D/3D initialization. (right) interaction panel indicating control mode, and cursors indicating the current gesture classification (unknown, release, grip). The four primary panes control orientation and translation. The bottom left pane and left sidebar indicate the operator's hand location in the Kinect's workspace. The top left pane toggles between the x-ray images.

location is controlled when the user is performing a release gesture.

Volume pose is manipulated using one x-ray image at a time. This allows the user to perform interaction in an intuitive manner. Intuitively it seems that it would be easier to interact with the volume itself and observe all renderings changing concurrently as the volume is manipulated. This is true when the interaction mode is based on dials manipulated with a mouse but is not the case when using hand gestures. By performing the pose manipulations with respect to the coordinate axes of the current reference x-ray, hand motion in the virtual and physical worlds correspond. That is, moving ones hand to the left indicates that the volume should move to the left of the current image. If we utilized all images concurrently and interpreted the hand motion as motion with respect to the volume's coordinate system, translation in a certain direction in the physical world corresponds to an arbitrary translation in the virtual world, resulting in a more complex interaction.

Control of the volume's orientation is performed independently for each of the rotation axes. Once the appropriate rotation pane is "gripped," back and forth motion of the hand controls the volume's rotation. Control of the volume's location is performed using combined in plane translation. Once this pane is "gripped" translation in the XY plane is used to modify the volume's location.

At first glance, it appears that this interface provides five degrees of freedom and thus cannot enable the operator to completely control the volume's pose, which has six degrees of freedom. In practice, this is not the case due to our manipulation approach which is done with respect to the x-ray coordinate system. The operator actually has $5n$ degrees of freedom to manipulate the volume, where n is the number of x-ray images. The only situation where this is an issue is when using a single image for registration. This case is not relevant for our target applications, but can easily be accommodated by adding the third, out of plane translation, to the translation pane.

Given that our pose manipulations are done with respect to the current x-ray coordinate system, a user interaction yields a *virtual* pose update to the x-ray camera, and the equivalent *actual* pose update to the 3D data needs to be computed accordingly. This is essential in order to achieve correct initialization, while providing the effect of *what you see is what you operate*. Figure 3 shows the relationship and conversion between these two types of pose updates.

A user interaction produces $T_c^{c'}$. Since $T_{s'}^{c} = T_s^{c'}$, and $T_{s'}^{c} = T_s^c T_{s'}^s$, $T_s^{c'} = T_c^{c'} T_s^c$, then the pose update to the 3D data is obtained by

$$T_{s'}^{s} = (T_s^c)^{-1} T_c^{c'} T_s^c, \tag{1}$$

$$= (T_s^w)^{-1} T_c^w T_c^{c'} (T_c^w)^{-1} T_s^w, \tag{2}$$

where T_c^w and T_s^w are the current poses of the x-ray and 3D data in the world coordinate frame, respectively. The new pose of the 3D data is obtained as

$$T_{s'}^{w} = T_s^w T_{s'}^s. \tag{3}$$

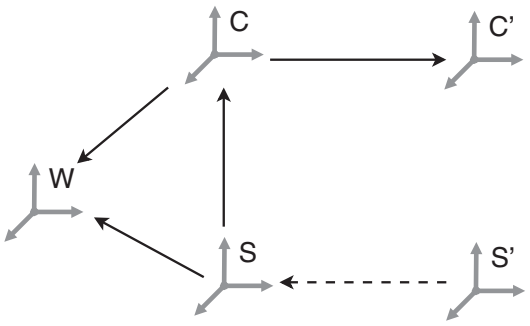

FIG. 3. Calculation of pose update to the 3D data for a user interaction. W is the world coordinate frame, C and C' are the coordinate frames of the reference x-ray at its current and *virtual* new positions, S and S' are the coordinate frames of the 3D data at its current and *actual* new positions. Dashed line marks the unknown transformation.

2.A.1.c. Gesture detection. We perform gesture detection using the skeletal model and depth data acquired by the Kinect sensor. The Kinect automatically identifies potential users and reports the joint positions of a skeletal model aligned with each of them. In our case, we select the user closest to the Kinect device. We use the Kinect in what is known as "seated mode."[16] This operating mode is optimized for detecting and tracking the skeletal joints above the hip. This is more appropriate for our application as we expect the surgeon to be standing behind the operating table.

Hand gestures are detected and classified using a preprocessed subregion of the depth data which is obtained as follows:

1. Region of interest detection: Given that we are only interested in right hand gestures we obtain the right hand joint position in the Kinect workspace. We then project this position onto the depth image and select a 140×140 pixel sized window centered on this position. This specific window size was determined empirically, such that it is sufficiently large to encompass an open hand throughout the interaction space of the Kinect.
2. Subsampling and depth modification: We downsample the region of interest by a factor of two, and discard depth values that are ± 10 cm from the hand position by setting them to zero. We then normalize the depth map to be in [0,1] so that the resulting map is independent of the users location in the workspace. The resulting 70×70 normalized depth map is the input for our classifier.

Figure 4 illustrates the data acquisition process described above.

To control the interaction panel, we defined two gestures: grip (fist) and release (open hand). Additionally, we defined a third "neutral gesture" class which consists of a variety of hand configurations which are not clear fist or open hand. All gestures classified as neutral are ignored by the interaction panel.

We use supervised multi-class classification to detect gestures, using a support vector machine (SVM) framework.[17] As

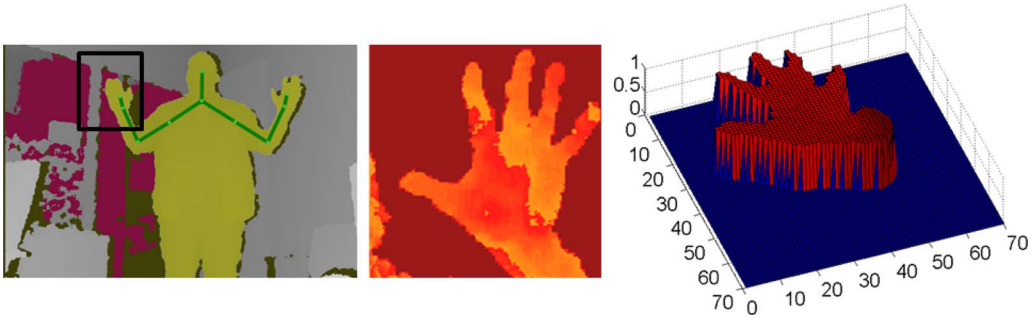

FIG. 4. Data acquisition steps shown from left to right: Original depth map where lighter colors are closer and darker colors are farther away. The joint positions are visualized as a skeleton overlaid onto the depth map; ROI surrounding the right hand; and downsampled and normalized depth map in ROI.

the SVM is a binary classifier we used the "one-against-one" approach to perform multi-class classification.[18]

The SVM solves the following optimization problem. Given training vectors $\mathbf{x_i} \in R^n, i = 1, \ldots, l$, in two classes and a label vector $\mathbf{y} \in R^l$ such that $y_i \in \{1, -1\}$,

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i,$$

$$\text{subject to } y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, i = 1, \ldots, l,$$

where $\phi(\mathbf{x}_i)$ maps $\mathbf{x}_i$ into a higher-dimensional space and $C > 0$ is the regularization parameter. We used the Gaussian radial basis function (RBF) kernel as our mapping function:

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}. \quad (4)$$

We have two free parameter values which we need to set, C and γ. In our case, we are using the SVM implementation of LIBSVM,[18] and have empirically found the default values, $C = 1$ and $\gamma = 1/\#features$, give satisfactory results.

To err on the conservative side, we have less confidence in the classification of a feature vector which is found to be close to the decision boundary. We thus use a threshold on this distance and accept the results of our classifier only if the feature vector is far from the boundary. When this is not the case, we ignore the current classification and set it to the previous one.

To assess the appropriateness of our choice to use the 4900 features obtained from the depth map, data were acquired from one participant performing different gestures and labeled as release, grip, or neutral. We then used the original data as input for a k-means clustering algorithm. We observed that a naive clustering on the data gives 87.5% accuracy (5779 of 6600 gestures) which indicates that our classes are well defined in this parameter space.

To train and test the classifier, five people with different hand sizes performed the specified gestures within the interaction space of the Kinect sensor at four locations in the workspace. Figure 5 illustrates the variability associated with these users. In total, we collected 12000 gestures to train and test the classifier.

We used a leave-one-out cross validation to test our classifier, where we used four of the subjects to train the classifier and one for testing in each turn. Classifier accuracies were 89%, 93%, 92%, 94%, and 98%. This classification rate is sufficiently accurate for our purposes.

2.A.2. AR-based initialization

In procedures that already utilize a tracking system, we propose an AR based approach to initialization, using a tracked tool to control the volume rendering overlay. Initialization is performed in two steps: pose planning in the virtual world, and interaction in the physical world. In the pose planning step, the user positions a model of the physical tool next to the volumetric representation of the anatomical structure obtained from CT or MR. In the interaction step, the user mimics the planned tool pose in the physical world. That is, they attempt to position the tool in the same pose relative to the anatomy as was done in the virtual world. Figure 6 illustrates the concept and coordinate systems involved in this approach.

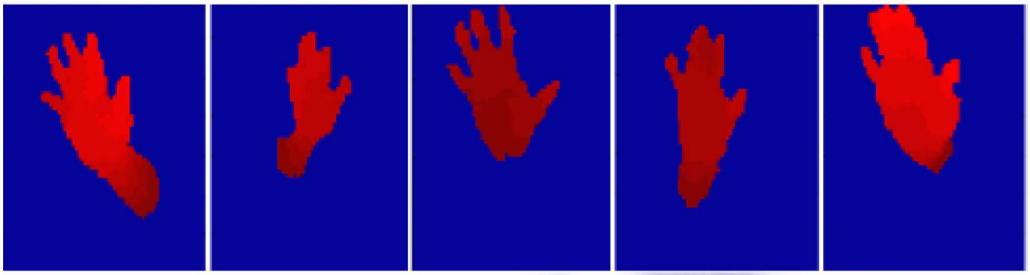

FIG. 5. Variability of hand sizes used for training the classifier.

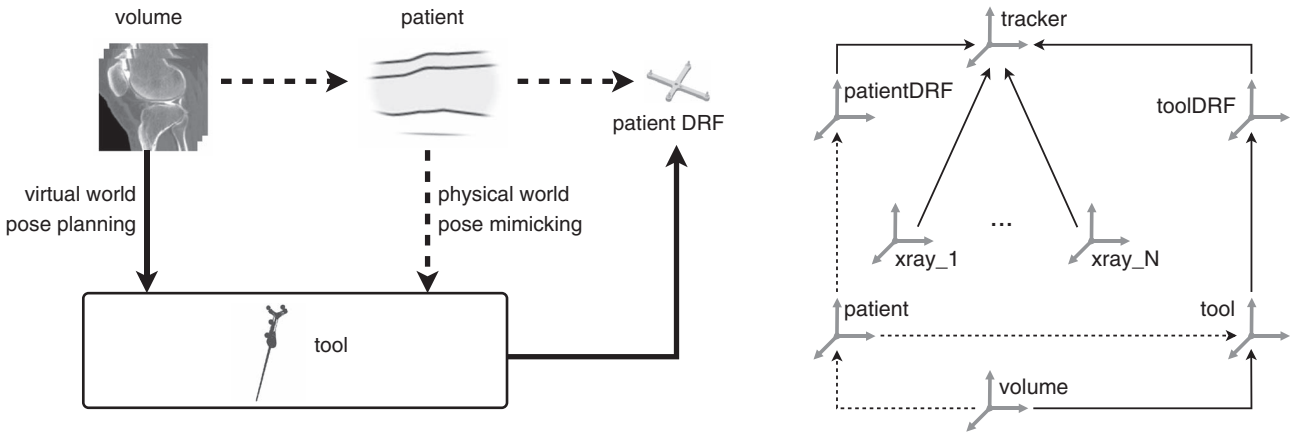

FIG. 6. (a) The user attempts to mimic the volume-tool pose defined in the virtual world to the patient-tool pose in the physical world. That is, positioning the tool in the physical world such that $T^{\text{tool}}_{\text{volume}} = T^{\text{tool}}_{\text{patient}}$. When this happens, the volume and patient coordinate frames coincide, and the unknown $T^{\text{patientDRF}}_{\text{patient}}$ can be computed by composing the known $T^{\text{tool}}_{\text{volume}}$ with the tracked $T^{\text{patientDRF}}_{\text{tool}}$. (b) Coordinate systems involved in the initialization, solid lines denote known transformations, dashed denote transformations to be determined.

The goal of initialization is to estimate the transformation from the patient's coordinate system to the patient's dynamic reference frame (DRF), $T^{\text{patientDRF}}_{\text{patient}}$. The following transformations are known: $T^{\text{tracker}}_{\text{patientDRF}}$ and $T^{\text{tracker}}_{\text{toolDRF}}$ are given by the tracking system; $T^{\text{toolDRF}}_{\text{tool}}$ is obtained from tool calibration; $T_{x-\text{ray}_i}$ are obtained from camera calibration and tracking; and $T^{\text{tool}}_{\text{volume}}$ is obtained from the planning step.

The transformation we want to estimate is given by:

$$T^{\text{patientDRF}}_{\text{patient}} = \left(T^{\text{tracker}}_{\text{patientDRF}}\right)^{-1} T^{\text{tracker}}_{\text{toolDRF}} T^{\text{toolDRF}}_{\text{tool}} T^{\text{tool}}_{\text{patient}}$$

$$\Downarrow$$

$$T^{\text{patientDRF}}_{\text{patient}} = T^{\text{patientDRF}}_{\text{tool}} T^{\text{tool}}_{\text{patient}}.$$

If the planned transformation is mimicked accurately in the OR, we have $T^{\text{tool}}_{\text{volume}} = T^{\text{tool}}_{\text{patient}}$, which gives us the desired transformation by substitution into the previous equation to yield.

$$T^{\text{patientDRF}}_{\text{patient}} = T^{\text{patientDRF}}_{\text{tool}} T^{\text{toolmodel}}_{\text{volume}}. \quad (5)$$

Depending on the accuracy requirements imposed by the subsequent registration algorithm, pose mimicking can be performed using one of two ways: instant coarse initialization, and interactive AR-based initialization. The former is applicable for a variety of intraoperative registration methods as it does not utilize any intraoperative images. The later does require availability of intraoperative images and is thus only applicable to procedures where intraoperative imaging is used.

2.A.2.a. Preoperative planning. The goal of planning is to define the relative pose between the *tool* and the *preoperative volume*. This is performed using a graphical user interface [Fig. 7(a)], within which the poses of the volume and the tool can be manipulated individually or concurrently. The tool is represented using a mesh model, which also defines the coordinate frame of the tool, and its pose should match its intended pose in the physical world. It is the user's responsibility to position the tool in a valid location, which means that the tool

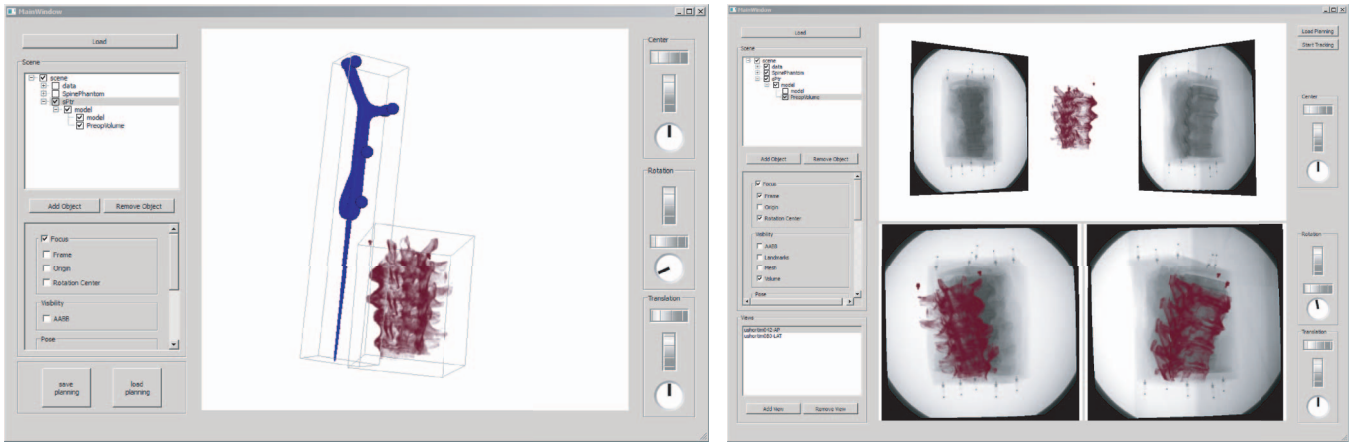

FIG. 7. Graphical user interfaces for (a) pose planning in virtual world and (b) pose mimicking in physical world.

cannot overlap with anatomical structures and, if using an optical tracking system, its planned position is expected to be visible in the OR.

To obtain an informative volume rendering in the subsequent initialization step, the transfer function for volume rendering is also interactively chosen in this stage. We use the same approach for registration of CT and MR. Given that in x-ray images the visible structures are primarily bones, we require the user to manipulate the transfer function so that these structures are visible to them. This does not imply that the transfer function is optimal, only that for the specific user it yields a visually clear set of anatomical structures.

It should be noted that our approach imposes several requirements on the design of the tracked tool. It must provide six degrees of freedom so that we can manipulate the volume pose in the physical world, and it should not be symmetrical so that the user can visually distinguish between different tool poses. That is, a cylindrical tool such as a needle is best avoided as it defines an infinite number of poses which only differ in rotation about the needle axis. One can design a specific tool based on these requirements, but this is most often not necessary. In our case, we utilize a pointer probe which is available as part of the navigation system.

2.A.2.b. Instant, coarse, initialization. The instant, coarse, initialization approach consists of a single step. The planned tool pose is mimicked in the OR by placing the tool besides the patient as planned and initiating the initialization with a foot switch. No further user interaction is required, and the desired transformations are estimated instantly. Obviously the accuracy of the result depends on the difference between $T^{\text{tool}}_{\text{volume}}$ and $T^{\text{tool}}_{\text{patient}}$, which introduces errors in computations of both $T^{\text{patient}}_{\text{volume}}$ and $T^{\text{patientDRF}}_{\text{patient}}$.

While this method is simple and fast, the initialization accuracy depends on how accurately the planned transformation can be replicated in the OR. By using anatomical landmarks, one can plan easy-to-reproduce poses which can provide relatively accurate initializations. As no intraoperative modality is involved, the method can potentially be used to initialize other forms of registration (e.g., point cloud/surface).

2.A.2.c. AR-based initialization. The interactive AR-based initialization approach is based on visual feedback which allows the operator to iteratively steer the volume toward its correct pose. We achieve this by real-time direct volume rendering which is overlaid onto the x-ray images. In our case, we perform hardware accelerated volume rendering in parallel for 2-3 images [Fig. 7(b)]. It should be noted that the camera parameters used to perform the rendering are specific to each x-ray image and are obtained from accurate calibration of the clinical imaging system.

To use the AR-based approach, the user starts by performing the coarse initialization as described above. This is required so that there is a reasonable overlap between the rendered images and the x-rays. Then, the user translates and rotates the tracked tool based on the AR views with the goal of maximizing the visual similarity between the overlaid volume rendering and underlying x-ray images. The process continues until a good overall overlay between the x-rays and the corresponding renderings is achieved. The maximal overlap is obtained when $T^{\text{tool}}_{\text{volume}} = T^{\text{tool}}_{\text{patient}}$, and the desired transformation is computed as described above. Again, the accuracy of the result depends on the difference between these two transformations.

2.B. Experimental evaluation

We evaluate our initialization approaches using three publicly available reference data sets for 2D/3D registration. The first data set[19] is from the Image Science Institute (ISI), Netherlands, and consists of images from a spine phantom containing three vertebrae. The second data set[20] is from the University of Ljubljana, from a phantom consisting of five lumbar vertebrae. The third data set[21] is from the Medical University of Vienna, and consists of a cadaver animal head. Unlike the previous two data sets, this data set contains a significant amount of soft tissue which is visible in the x-ray images.

From each of the data sets, we used two x-ray images, one MR, one CT, and the reference transformations for the MR and CT. The reference transformations position the volumes in the "tracker", common, coordinate frame to match the corresponding x-ray images. Figure 8 shows all x-ray images used in this study, as well as the corresponding volume renderings of CT and MR at the reference poses.

Initialization errors were evaluated using a set of targets within the 3D data (MR or CT). To obtain the targets, the 3D data were interactively thresholded such that most of the selected points were from the bone, then edge points were extracted from the thresholded volume and used as the targets. Let T_g and T_r be the transformations of a 3D data S at its ground truth position g and initialized position r, respectively. The initialization error is calculated using mean target registration error (mTRE), which is defined as[19]

$$mTRE(r;S) = \frac{1}{N}\sum_{i=1}^{N} \|T_r p_i - T_g p_i\|, \tag{6}$$

where p_i is a target at index i, r are the initialization transformation parameters, and N is the total number of targets.

For each reference data set, ten testing cases were created and were used for both x-ray/MR and x-ray/CT experiments. Three operators, two developers of the algorithms (P1, P2), and one physician (P3) performed the experiments using both the gesture based and AR based approaches. Initial and final mTREs as well as interaction time were recorded.

2.B.1. Gesture-based initialization

For each data set and each of the MR and CT images, we randomly perturbed the six degree-of-freedom reference transformation to generate ten testing cases. The perturbation range was $\pm 90^\circ$ (around the center of the MR or CT) for the rotation components and ± 50 mm for the translation components.

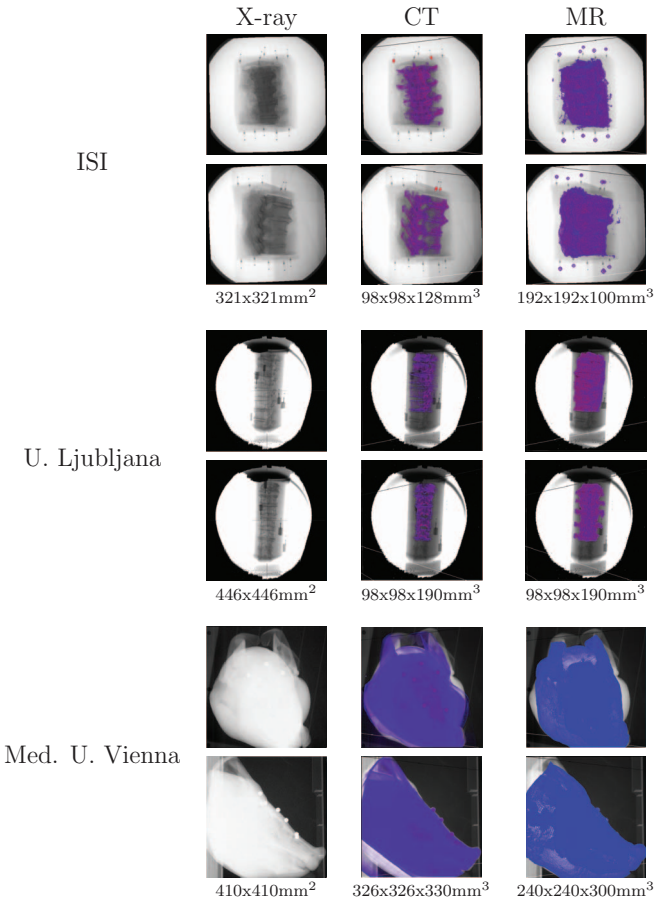

FIG. 8. Reference data sets used in this study. The x-ray images for the ISI and Vienna data sets were provided in the anterior-posterior and lateral views. The Ljubljana data set provided 18 evenly spaced x-ray images around the spinal cord, and we selected two images (000 and 006) in our study. Original x-ray images for this data set had low contrast, so they were thresholded and enhanced to provide better visualization. The right two columns show the corresponding 2D rendering of the CT and MR images at their ground truth positions. They show the differences between x-ray images and rendering of the 3D data. The physical dimensions for x-rays, CT, and MR are shown underneath the images for each data set.

To accurately position the volume, we assume that the user is familiar with the anatomical structure. This assumption is valid for clinicians but was not valid for all participants in this experiment. To address this issue, we allowed each participant to spend 10 min studying the anatomy on screen before starting the experiments. During this period, the participant arbitrarily manipulated the volumetric data in order to obtain an understanding of the spatial structure of the anatomy.

Note that the initial mTRE values reported in our results are the errors after automatic positioning of the volume (see Sec. 2.A.1) and not those obtained from perturbation.

2.B.2. AR-based initialization

In the reference data sets, the reference transformations are given with respect to their own "tracker's," common, coordinate frame. We need to link these "tracker" coordinate frames to our physical setup and tracking system. Figure 9(a) illustrates how the reference transformation is established, and how the error transformation is computed. $T^{\text{tracker}'}_{\text{patient}}$, $T^{\text{tracker}'}_{\text{x-ray\_ap}}$ and $T^{\text{tracker}'}_{\text{x-ray\_lat}}$ were provided as part of the reference data. The transformation we are interested in, $T^{\text{patientDRF}}_{\text{patient}}$, is unknown as we do not have the physical phantoms from which the reference data sets were created. We thus need to make an arbitrary choice, relating a physical, tracked, reference frame to the phantom. Once this transformation is established, we can compute the transformation $T^{\text{tracker}}_{\text{tracker}'}$ accordingly to obtain the reference transformations with respect to our tracker. Then the error transformation between the estimated and ground-truth poses is computed as

$$T^{\text{patient}}_{\text{volume}} = \left(T^{\text{tracker}}_{\text{patientDRF}} T^{\text{patientDRF}}_{\text{patient}}\right)^{-1} T^{\text{tracker}}_{\text{toolDRF}} T^{\text{toolDRF}}_{\text{tool}} T^{\text{tool}}_{\text{volume}}. \tag{7}$$

Note that when the user is able to exactly mimic the planned tool position in the physical world we have $T^{\text{patient}}_{\text{volume}} = I$.

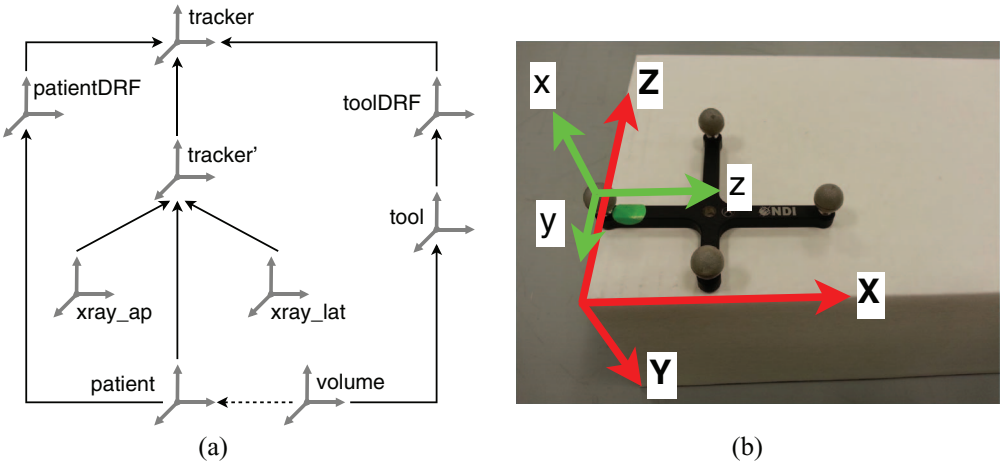

FIG. 9. (a) Transformations involved in the validation of AR-based initialization approach. The tracker' coordinate system is the common/world coordinate system used by the reference data set. (b) Definition of reference transformation. The transformation between the patient, bottom, and patient DRF, top, coordinate systems was used as the reference for our experiments.

In our experiments, $T_{\text{patient}}^{\text{patientDRF}}$ was chosen based on the bounding box of the volume. First, a cardboard box, Fig. 9(b), was used to represent the physical patient. The box roughly matches the volume's bounding box in size, and its coordinate frame is aligned with the volume's coordinate frame. Then the patient DRF was placed at the lower-left corner of the xz-surface of the box. Finally, we obtained the coordinates of three known points on the box in the DRF's coordinate system. Thus we have the coordinates of the same points in the patient coordinate system and in the DRF coordinate system. From this setup, $T_{\text{patient}}^{\text{patientDRF}}$ is readily available via paired point rigid registration.[22]

We used the Polaris Vicra optical tracking system from Northern Digital Inc. (Waterloo, ON, Canada) to evaluate our approach.

For each data set, x-ray/MR and x-ray/CT, ten pose plans were created (see Sec. 2.A.2.a), and the experiments were performed using both the instant, coarse, and interactive, AR-based, initialization approaches.

3. RESULTS

3.A. Gesture-based initialization

Figures 10 and 11 summarize the experimental results for the gesture based interactive initialization.

For x-ray/MR initialization, we observed an accuracy of, 9.3 ± 5.0 mm, versus an accuracy of 7.4 ± 5.0 mm, for x-ray/CT. This difference is statistically significant (Wilcoxon Rank-Sum test, $p = 0.0011$).

For x-ray/MR initialization we observed longer interaction times, 146.3 ± 73.0 s, versus an interaction time of 132.1 ± 66.4 s, for x-ray/CT. This difference is not statistically significant (Wilcoxon Rank-Sum test, $p = 0.11$).

The overall initialization error including all participants, modalities and data sets were 8.35 ± 5.0 mm within interaction times of 139.2 ± 70.0 s. These results satisfy the requirements of most 2D/3D registration applications.

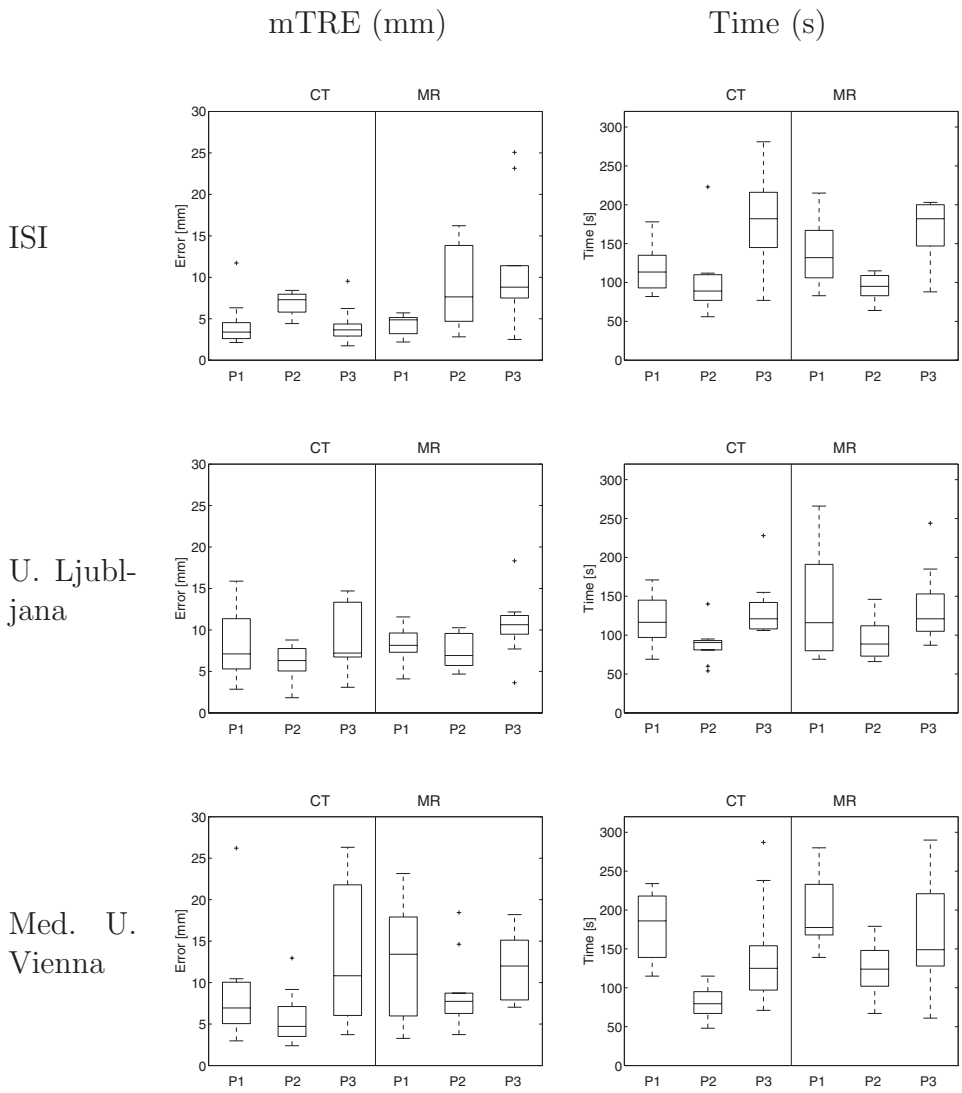

FIG. 10. Experimental results for gesture-based initialization. (left column) distribution of mTRE per operator, and (right column) distribution of interaction time per operator.

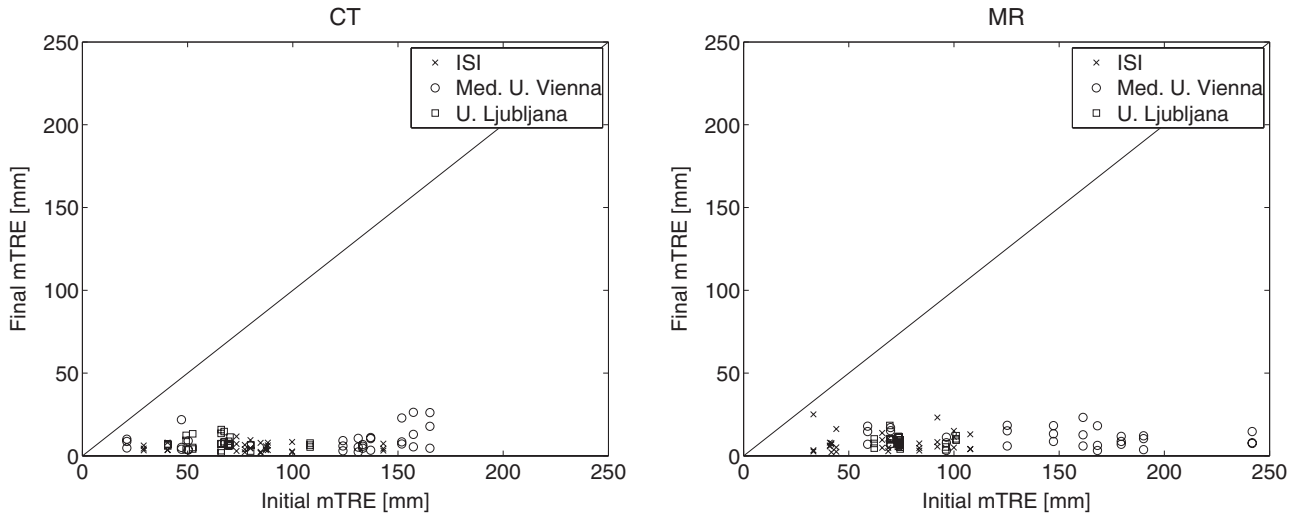

FIG. 11. Initial versus final mTRE for the gesture-based method, diagonal marks the no-change line. Note that the initial mTREs for the Vienna data set (circle marker symbol) are larger due to the "lever" effect associated with rotating a larger object.

3.B. AR-based initialization

Table I summarizes the experimental results for the instant, coarse, initialization. We can see that, as the only information to guide the pose mimicking actions was the planned poses in the user's mind, the method resulted in a relatively high average mTRE (28-40 mm). However, it should be noted that the rotational errors are relatively low, while the translational errors are high. This fact can benefit registration algorithms as correcting rotation errors is more challenging than correcting translation errors. Also, we observed that, for each data set, the rotational errors are more dominant along one axis. Not surprisingly this axis corresponds to the long axis of our tool. As we noted in Sec. 2.A.2.a, the uncertainty in orientation when using a cylindrical like tool is higher around its main axis. This theoretical observation is reflected in practice by our results.

Figure 12 presents the experimental results for the interactive AR based initialization, and Table II reports the consolidated initialization errors and interaction times according to different classification categories. For the three participants, no obvious differences in mTRE or completion times were observed. This suggests that the proposed approach is equally intuitive for an operator that had no prior experience working with a tracking system (P3), and for operators that have worked with tracking systems in the past (P1 and P2).

TABLE I. Experimental results for instant coarse initialization (for each reference data set, the results were summarized from all participants and both x-ray/CT and x-ray/MR experiments).

	ISI	U. Ljubljana	Med. U. Vienna
mTRE (mm)	27.5 ± 13.7	28.0 ± 14.7	39.7 ± 19.2
θ_x (°)	0.8 ± 3.1	0.1 ± 2.9	1.6 ± 8.7
θ_y (°)	-4.2 ± 6.5	-0.2 ± 2.0	1.4 ± 2.1
θ_z (°)	1.7 ± 2.3	-5.5 ± 4.7	-0.1 ± 3.0
t_x (mm)	-11.8 ± 11.7	-18.7 ± 18.6	-7.0 ± 14.3
t_y (mm)	9.3 ± 11.3	7.8 ± 14.5	3.8 ± 17.6
t_z (mm)	-6.0 ± 16.4	5.4 ± 15.4	6.5 ± 10.3

For x-ray/MR initialization, we observed an accuracy of, 7.2 ± 3.2 mm, versus an accuracy of 8.3 ± 5.0 mm, for x-ray/CT. This difference is not statistically significant (Wilcoxon Rank-Sum test, $p = 0.4501$). The smaller initialization errors are mainly due to the Ljubljana data (see Fig. 12, the second row of the mTRE plots), whose MR rendering actually provided more useful structures to overlay with the x-ray images. The difference in mTREs for each of the data sets is mainly associated with its physical dimensions. A large physical dimension yielded large mTREs, e.g., the Vienna data, and small physical dimensions yielded smaller mTREs, e.g., the ISI data. This is reasonable as small orientation errors can yield large registration errors for data sets with larger dimensions.

For x-ray/MR initialization, we observed shorter interaction times, 44 ± 32 s, versus an interaction time of 58 ± 52 s, for x-ray/CT. This difference is not statistically significant (Wilcoxon Rank-Sum test, $p = 0.0627$). This was observed for all three operators. These results would seem to be counterintuitive. Our expectation was that registration to CT would be faster than registration to MR, given the better visual quality of the volume rendering. In practice, the improved visual quality actually leads to longer interaction times, with the operator tempted to invest a long time in minor accuracy improvements. We also observed difference in interaction times between the data sets, most likely these are due to the different anatomical structures visible in the data sets.

Summarizing the results from all participants, modalities, and data sets, initialization errors of 7.7 ± 4.2 mm within interaction times of 51 ± 43 s were achieved. These numbers satisfy the requirements of most 2D/3D registration applications.

Figure 13 top row shows the relationship between the instant, coarse, initialization error, and the interactive AR based initialization error. We can see that, the interactive initialization improved the coarse alignments for all cases, and the final

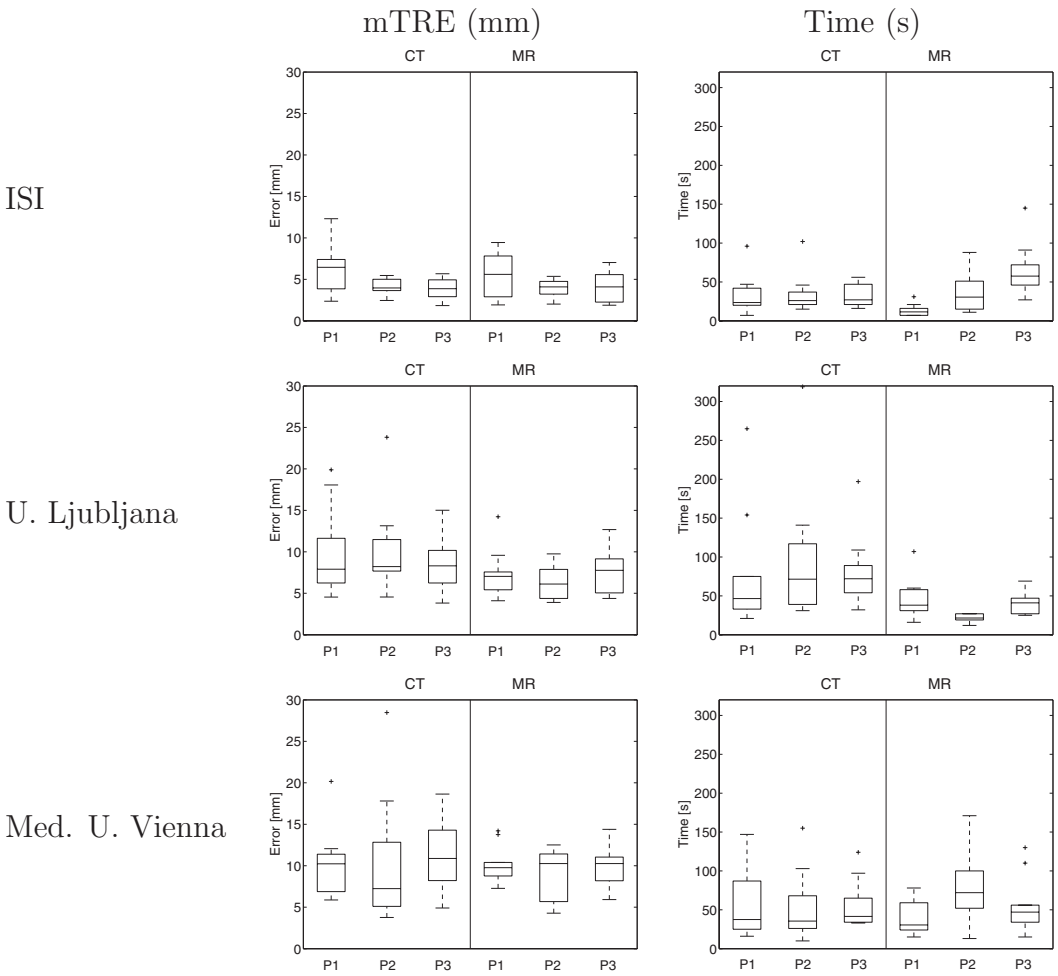

FIG. 12. Experimental results for interactive AR-based initialization. (left column) distribution of mTRE per operator, and (right column) distribution of interaction time per operator.

mTREs did not depend on the initial mTREs obtained by the coarse initialization. However, from Fig. 13 bottom row, we can see that the coarse mTREs did effect the interaction times.

During our experiments, we encountered several cases (<10) in which, the operator could not complete the initialization. It was visually clear that the volume was not aligned correctly but the operator was unable to manipulate the tool to improve the alignment. These cases have been excluded from the results reported above. We identified that these failures are due to a disconnect between hand-eye coordination arising from the users focus on screen while manipulating the tool without observing it. During the process the operator drifted considerably from their initial coarse setup without realizing it. This form of hand-eye disconnect is a well known phenomena in image-guided procedures, with several *in situ* display methods proposed to mitigate it.[23,24] In our case, this is easily mitigated, the operator need only repeat the coarse initialization if they realize they can no longer manipulate the AR view in an intuitive manner.

TABLE II. Consolidated results for interactive AR-based initialization (results were summarized according to different categories).

		mTRE (mm)	Time (s)
Participants	P1	8.2 ± 3.9	44 ± 43
	P2	7.3 ± 4.8	53 ± 51
	P3	7.6 ± 3.9	55 ± 34
Modalities	X-ray/CT	8.3 ± 5.0	58 ± 52
	X-ray/MR	7.2 ± 3.2	44 ± 32
Data Sets	ISI	4.7 ± 2.1	36 ± 27
	U. Ljubjana	8.3 ± 3.9	60 ± 56
	Med. U. Vienna	10.2 ± 4.2	56 ± 38
Overall		7.7 ± 4.2	51 ± 43

4. DISCUSSION AND CONCLUSIONS

We are investigating the use of diagnostic MR or CT images to guide orthopaedic surgical procedures in pediatrics, with a focus on using MR to guide anterior cruciate ligament (ACL) reconstruction surgery. To register the 3D image to the intraoperative setting, we propose to use 2D/3D anatomy based rigid registration. This requires initialization of the registration algorithm. In our case, none of the existing initialization algorithms is directly applicable, as discussed below.

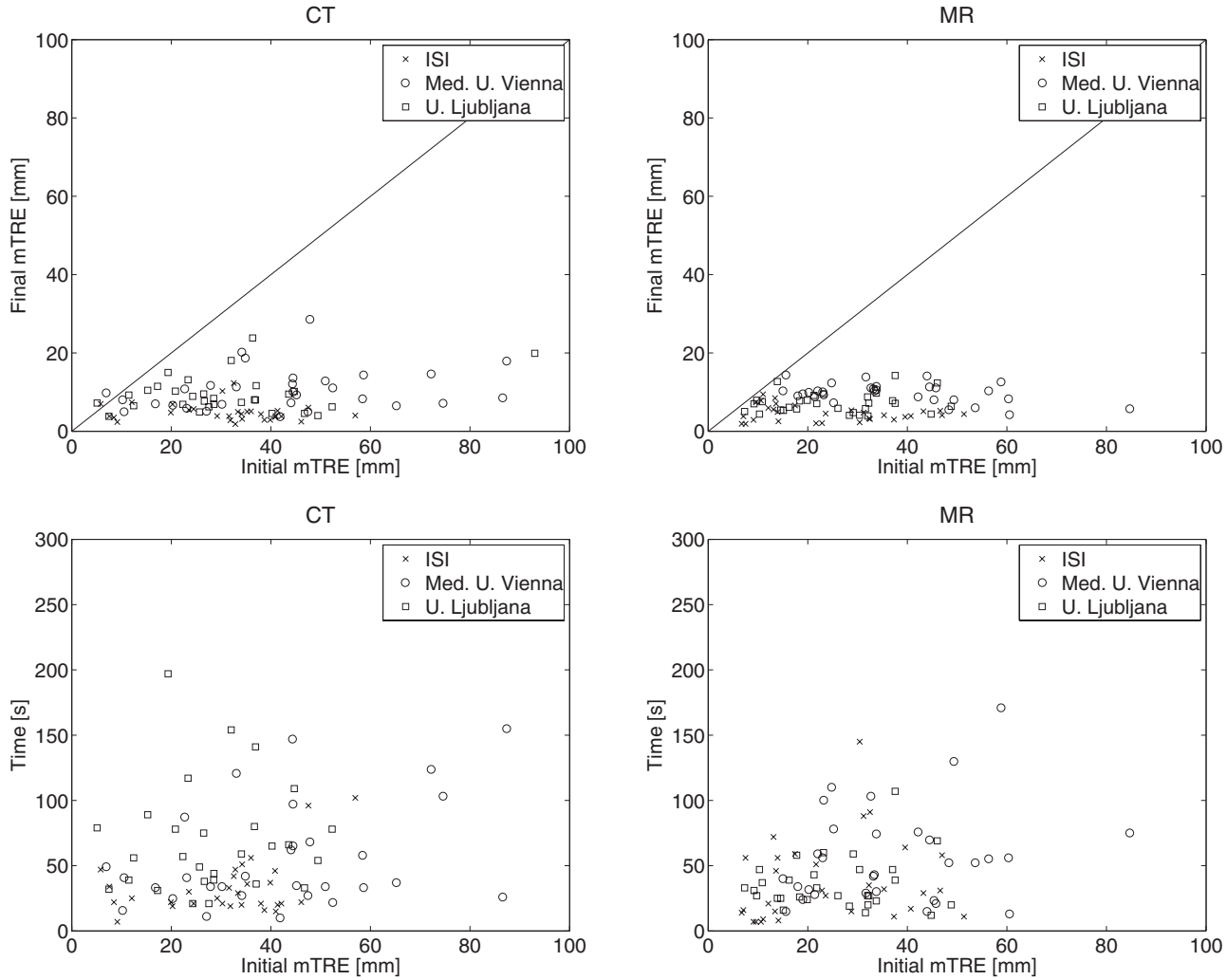

FIG. 13. Relationship between instant and interactive initialization: instant mTRE versus interactive mTRE (top), and instant mTRE versus interaction time (bottom).

Standard manual initialization using a keyboard and mouse is not appropriate for the operating room setting. This is due to the surgeon's desire to control the process, while dealing with the requirement for a sterile environment and the cramped quarters near the patient bed.

Coarse paired point registration requires that skin adhesive fiducials be placed on the patient prior to imaging or the ability to localize three or more anatomical landmarks. Use of fiducials is not applicable in many cases as it modifies the existing clinical workflow in which the diagnostic images are acquired long before the intervention. When using anatomical landmarks, it is often the case that there are not enough landmarks that can be easily and reliably localized. This is due to lack of physical access in minimally invasive surgery, or to difficulties identifying the same landmark in multiple x-rays.

Use of prior knowledge about the clinical setup is often not sufficiently accurate. This is primarily of concern in procedures where the pose of the anatomy of interest can have high variability. This is the case with ACL reconstruction surgery. The spatial relationships between the anatomical structures of interest, femur and tibia, and imaging equipment can vary considerably as the leg is manually positioned. The clinical setup for this procedure is shown in Fig. 14. The spatial relationship between the femur and tibia is dependent upon knee flexion and also varies considerably from the fully extended position in which the diagnostic MR is acquired.

In other surgical procedures, where the clinical setup does provide a sufficiently accurate initialization for some parameters but not others, our methods can be readily used to recover the missing information with the operator only using a subset (<6) of the degrees of freedom available to them. It should be noted that even for these settings the ability to manipulate all six degrees of freedom is important, as it provides an alternative initialization method when the clinical setup varies from the expected one, or when there are errors in the data (e.g., volumetric scan indicates Head First Supine when in practice it is Feet First Supine). In such cases, using all six degrees of freedom to perform initialization will enable the procedure to continue as planned, rather then abort the use of navigation guidance.

The brute force, coarse grid based search of the parameter space is also not applicable, as it implicitly relies on a

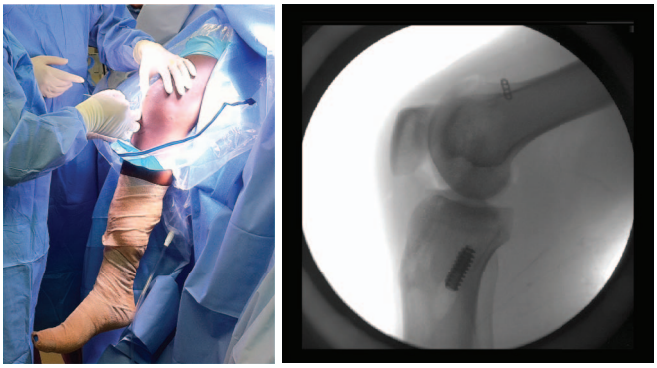

FIG. 14. Clinical setup for ACL reconstruction surgery, and a lateral x-ray (not from same procedure). X-ray shows knee flexed at about 120°. Preoperative diagnostic MR is acquired at full extension, 0° between femur and tibia.

transformation estimate obtained from the clinical setup. The search grid is then defined relative to these transformation parameter values.

The use of CBCT/CT or CBCT/MR intensity based registration is only applicable when an intraoperative CBCT system is available, which is not the current standard in most operating rooms. In addition, the acquisition of an intraoperative CBCT scan increases the radiation exposure to the patient. In the pediatric domain, this is of particular concern and is best avoided.[25]

The Fourier slice theorem based approach is only applicable for x-ray/CT registration, and its performance will degrade if the spatial relationship between the imaged anatomical structure changes or if there are additional objects in the x-ray which were not present in the CT. This is the rule and not the exception in orthopaedic surgery. In joint or spine procedures, we are dealing with multiple anatomical structures whose spatial relationship can change between preoperative imaging and the intraoperative setting. In addition, it is not uncommon to have medical apparatus visible in the x-ray images which was not part of the preoperative CT scan.

The virtual fiducial marker based approach utilizes a bootstrap type of methodology where the virtual fiducial marker coordinates are obtained in the CT coordinate system by performing 2D/3D registration based on anatomical landmarks. Subsequent 2D/3D registrations are based on the use of the virtual marker and knowledge of the clinical setup, which according to their experience provides orientation information which has an average reproducibility of $3^\circ-4^\circ$. This approach is not directly applicable in our setting as we do not deal with reinitialization, and thus this approach becomes equivalent to the coarse paired point registration method.

Our work was inspired by an observation made by Udupa *et al.*[26] in the context of segmentation which is equally relevant for registration, humans are highly adept at recognition. That is, we are extremely proficient at determining the presence and rough location of an object of interest in an image, and can do so in a robust manner in the presence of noise and overlapping structures. In the context of registration, these abilities allow an operator to robustly register a preoperative volume, MR, or CT to intraoperative x-rays based on visual similarity. The registration may not necessarily be accurate enough for navigation, but based on our results it is sufficiently accurate for initialization of an automated registration algorithm.

We also realize that surgeons prefer to directly control the systems at play in the operating room, and that control via proxy may be less than ideal. This was observed by Grätzel *et al.*[27] in the context of image-guided navigation. A surgeon provided instructions with the intent of performing a single mouse click, eventually taking 7 min to completion. As a consequence, we have developed two interactive methods for initialization of 2D/3D registration that are directly controlled by the surgeon. These methods were designed with integration into the clinical setting in mind. For procedures that currently do not use a navigation system, we introduced a gesture based interaction approach using the Microsoft Kinect. For procedures that do use a navigation system, we use the existing tracking system and introduce an AR-based approach.

The gesture based initialization approach only requires the introduction of a low cost piece of hardware into the interventional setting. Previous clinical applications of gesture based interaction using depth cameras such as the Kinect have primarily focused on browsing and interaction with volumetric images.[28–30] In that context, the use of gestures is geared toward completing a qualitative task. That is, the operator rotates the image until they are satisfied with the display, at which point the task is considered successful. In our case, the operator rotates the image until they are satisfied, but success is quantified in an objective manner based on the known reference transformations.

The AR-based initialization approach is designed with image-guided navigation procedures in mind. In this case, we do not introduce any additional equipment and utilize the existing tracking system. The user controls the pose of the volume using a tracked tool and visual feedback from the augmented view. A similar use of AR as a guidance approach is described in Ref. 31. In that work AR is used to provide acquisition guidance for freehand SPECT. A tracked Gamma probe acquires readings for the reconstruction, with the AR view guiding the physician to position the tracked probe in regions where there is not enough information to perform the reconstruction.

We evaluated both of our initialization approaches using three publicly available reference data sets. For the gesture based approach we obtained registration results with mTRE 8.35 ± 5.0 mm within 139.2 ± 70.0 s. For the AR-based approach we obtained registration results with mTRE 7.7 ± 4.2 in 51 ± 53 s.

We note that the registration errors associated with the Vienna data set are slightly larger than the other two data sets. This is primarily due to the soft tissue content visible in the x-ray images (see first column of Fig. 8). In x-ray images for ACL reconstruction surgery, our clinical application, we see more soft tissue than in the spine data sets, but much less than in the Vienna animal model (see Fig. 14). We thus expect our accuracy to be slightly better than that obtained for the Vienna data set, as the features visible in our data are closer to those visible in the spine models.

Based on the obtained results, we conclude that our interactive initialization approaches are sufficiently accurate as an initialization step for most 2D/3D anatomy based registration algorithms.

ACKNOWLEDGMENTS

The authors would like to thank E. Wilson for his help creating the mesh model of the tracked tool.

[a]Author to whom correspondence should be addressed. Electronic mail: ZYaniv@childrensnational.org

[1]*Image-Guided Interventions Technology and Applications*, edited by T. Peters and K. Cleary (Springer, New York, NY, 2008).
[2]A. Raabe, R. Krishnan, R. Wolff, E. Hermann, M. Zimmermann, and V. Seifert, "Laser surface scanning for patient registration in intracranial image-guided surgery," Neurosurgery **50**, 797–801 (2002).
[3]Z. Yaniv, "Rigid registration," *Image-Guided Interventions: Technology and Applications*, edited by T. Peters and K. Cleary (Springer-Verlag, New York, NY, 2008), Chap. 6.
[4]P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, "A review of 3D/2D registration methods for image-guided interventions," Med. Image Anal. **16**, 642–661 (2012).
[5]Y. Dori, M. Sarmiento, A. C. Glatz, M. J. Gillespie, V. M. Jones, M. A. Harris, K. K. Whitehead, M. A. Fogel, and J. J. Rome, "X-Ray magnetic resonance fusion to internal markers and utility in congenital heart disease catheterization," Circle Cardiovasc. Imaging **4**, 415–424 (2011).
[6]H. Kang, D. M. Lovelock, E. D. Yorke, S. Kriminski, N. Lee, and H. I. Amols, "Accurate positioning for head and neck cancer patients using 2D and 3D image guidance," J. Appl. Clin. Med. Phys. **12**, 86–96 (2011).
[7]S. Miao, J. Lucas, and R. Liao, "Automatic pose initialization for accurate 2-D/3-D registration applied to abdominal aortic aneurysm endovascular repair," Proc. SPIE **8316**, 83160Q (2012).
[8]A. K. George, M. Sonmez, R. J. Lederman, and A. Z. Faranesh, "Robust automatic rigid registration of MRI and X-ray using external fiducial markers for XFM-guided interventional procedures," Med. Phys. **38**, 125–141 (2011).
[9]K. P. Gall, L. J. Verhey, and M. Wagner, "Computer-assisted positioning of radiotherapy patients using implanted radiopaque fiducials," Med. Phys. **20**, 1153–1159 (1993).
[10]R. H. Gong, O. Güler, and Z. Yaniv, "Interactive initialization for 2D/3D intra-operative registration using the microsoft kinect," Proc. SPIE **8671**, 867128 (2013).
[11]Y. Otake, M. Armand, R. S. Armiger, M. D. M. Kutzer, E. Basafa, P. Kazanzides, and R. H. Taylor, "Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: Incorporation of fiducial-based C-arm tracking and GPU-acceleration," IEEE Trans. Med. Imaging **31**, 948–962 (2012).
[12]Y. Otake, S. Schafer, J. W. Stayman, W. Zbijewski, G. Kleinszig, R. Graumann, A. J. Khann, and J. H. Siewerdsen, "Automatic localization of target vertebrae in spine surgery using fast CT-to-fluoroscopy (3D-2D) image registration," Proc. SPIE **8316**, 83160N (2012).
[13]C. Gendrin, H. Furtado, C. Weber, C. Bloch, M. Figl, S. A. Pawiro, H. Bergmann, M. Stock, G. Fichtinger, D. Georg, and W. Birkfellner, "Monitoring tumor motion by real time 2D/3D registration during radiotherapy," Radiother. Oncol. **102**, 274–280 (2012).
[14]M. J. van der Bom, L. W. Bartels, M. J. Gounis, R. Homan, J. Timmer, M. A. Viergever, and J. P. W. Pluim, "Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation," Med. Phys. **37**, 1884–1892 (2010).
[15]A. Varnavas, T. Carrell, and G. P. Penney, "Increasing the automation of a 2D-3D registration system," IEEE Trans. Med. Imaging **32**, 387–399 (2013).
[16]Microsoft Kinect SDK, see URL http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx.
[17]C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn. **20**, 273–297 (1995).
[18]C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011).
[19]E. B. van de Kraats, G. P. Penney, D. Tomaževič, T. van Walsum, and W. J. Niessen, "Standardized evaluation methodology for 2D-3D registration," IEEE Trans. Med. Imaging **24**, 1177–1189 (2005).
[20]D. Tomaževič, B. Likar, and F. Pernuš, "Gold standard data for evaluation and comparison of 3D/2D registration methods," Comput. Aided Surg. **9**, 137–144 (2004).
[21]S. A. Pawiro, P. Markelj, F. Pernus, C. Gendrin, M. Figl, C. Weber, F. Kainberger, I. Nobauer-Huhmann, H. Bergmeister, M. Stock, D. Georg, H. Bergmann, and W. Birkfellner, "Validation for 2D/3D registration I: A new gold standard data set," Med. Phys. **38**, 1481–1490 (2011).
[22]B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. Am. A **4**, 629–642 (1987).
[23]G. D. Stetten and V. Chib, "Overlaying ultrasound images on direct vision," J. Ultrasound Med. **20**, 235–240 (2001).
[24]G. Fichtinger, A. Deguet, K. Masamune, E. Balogh, G. S. Fischer, H. Mathieu, R. H. Taylor, S. J. Zinreich, and L. M. Fayad, "Image overlay guidance for needle insertion in CT scanner," IEEE Trans. Biomed. Eng. **52**, 1415–1424 (2005).
[25]M. S. Pearce *et al.*, "Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study," Lancet **380**, 499–505 (2012).
[26]J. K. Udupa *et al.*, "A framework for evaluating image segmentation algorithms," Comput. Med. Imaging Graph. **30**, 75–87 (2006).
[27]C. Grätzel, T. Fong, S. Grange, and C. Baur, "A non-contact mouse for surgeon-computer interaction," Technol. Health Care **12**, 245–257 (2004).
[28]L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross, "You can't touch this: Touch-free navigation through radiological images," Surg. Innov. **19**, 301–307 (2012).
[29]M. G. Jacob, J. P. Wachs, and R. A. Packer, "Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images," J. Am. Med. Inform. Assoc. **20**, e183–e186 (2013).
[30]G. C. S. Ruppert, L. O. Reis, P. H. J. Amorim, T. F. de Moraes, and J. V. L. da Silva, "Touchless gesture user interface for interactive image visualization in urological surgery," World J. Urol. **30**, 687–691 (2012).
[31]A. Okur, S.-A. Ahmadi, A. Bigdelou, T. Wendler, and N. Navab, "MR in OR: First analysis of AR/VR visualization in 100 intra-operative Freehand SPECT acquisitions," in *Proceedings of the 10th International Symposium on Mixed and Augmented Reality (ISMAR)*, Basel, Switzerland, October 26–29 (IEEE, 2011), pp. 211–218.