Ren Hui Gong, Oezgur Gueler, Ziv R. Yaniv, "Interactive initialization for 2D/3D intra-operative registration using the Microsoft Kinect", Proceedings of SPIE Vol. 8671, 867128 (2013).

http://dx.doi.org/10.1117/12.2008525

Interactive Initialization for 2D/3D Intra-Operative Registration using the Microsoft Kinect

Ren Hui Gong[a*], Özgur Güler[a*] and Ziv Yaniv[a]

[a]Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington, DC 20010, USA

ABSTRACT

All 2D/3D anatomy based rigid registration algorithms are iterative, requiring an initial estimate of the 3D data pose. Current initialization methods have limited applicability in the operating room setting, due to the constraints imposed by this environment or due to insufficient accuracy. In this work we use the Microsoft Kinect device to allow the surgeon to interactively initialize the registration process. A Kinect sensor is used to simulate the mouse-based operations in a conventional manual initialization approach, obviating the need for physical contact with an input device. Different gestures from both arms are detected from the sensor in order to set or switch the required working contexts. 3D hand motion provides the six degree-of-freedom controls for manipulating the pre-operative data in the 3D space. We evaluated our method for both X-ray/CT and X-ray/MR initialization using three publicly available reference data sets. Results show that, with initial target registration errors of 117.7 ± 28.9 mm a user is able to achieve final errors of 5.9 ± 2.6 mm within 158 ± 65 sec using the Kinect-based approach, compared to 4.8 ± 2.0 mm and 88 ± 60 sec when using the mouse for interaction. Based on these results we conclude that this method is sufficiently accurate for initialization of X-ray/CT and X-ray/MR registration in the OR.

Keywords: image-guided therapy, registration initialization, human computer interface, Microsoft Kinect

1. INTRODUCTION

The subject of 2D/3D, X-ray/CT, X-ray/MR and X-ray/atlas, anatomy based rigid registration has been studied extensively, resulting in a large number of algorithms, all of which are iterative.[1] Empirical evaluations using clinical data have shown registration algorithms successfully converge 95% of the time if after initialization the mean target registration error is on the order of 4-11 millimeters,[2] depending on the size and shape of the anatomical structure. As a consequence, clinically viable initialization approaches are required.

Currently, the most common initialization approaches include:[2,3] (1) Manual initialization - interactively input transformation parameter values using the keyboard or mouse. These are used to generate 2D images from the 3D data. The user actively searches for parameter values which result in a generated image that is visually similar to the medical one; (2) Clinical setup using the known geometry of the intra-operative imaging system to bound the transformation parameters; and (3) Coarse, approximate, registration - using a paired point analytic registration algorithm with point data obtained either from skin adhesive fiducials or anatomical landmarks. Another, less common, option arises when using an intra-operative Cone-Beam CT system, acquire an initial volume and perform 3D/3D registration. In the latter case 2D/3D registration is used to update the initial transformation.

We are investigating the use of MR and CT images for guiding pediatric orthopedic procedures. To register the 3D data to the intra-operative setting we propose to use 2D/3D anatomy based registration. In our case, none of the initialization algorithms is directly applicable. Manual initialization using keyboard and mouse in the operating room is less advisable, due to the requirements for a sterile environment and the surgeon's desire to control the process. Use of the known geometry of the imaging setup is often not sufficiently accurate, primarily when the patient position is not close to the expected one. Finally using analytic paired point registration

E-mail: {rgong,oguler,zyaniv}@cnmc.org.
* These authors contributed equally to this work.

Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by David R. Holmes III, Ziv R. Yaniv, Proc. of SPIE Vol. 8671, 867128
© 2013 SPIE ·CCC code: 1605-7422/13/$18 · doi: 10.1117/12.2008525

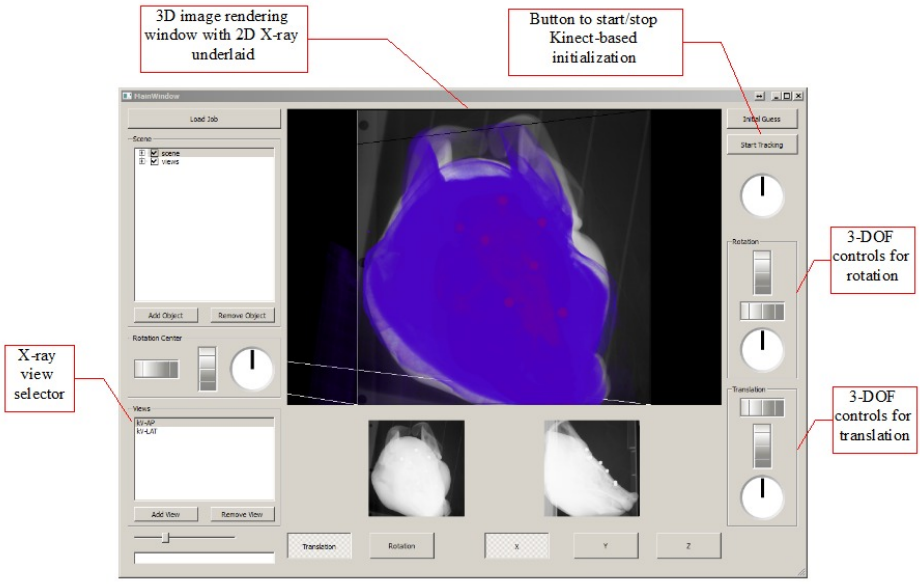

Figure 1. Graphical User Interface for 2D/3D initialization. The user can modify the pose of the volumetric data by using the mouse or via gesture based interaction using the Microsoft Kinect.

requires that the markers be placed prior to imaging, which is often not possible, or that there is physical access to anatomical landmarks which is not the case in many minimally invasive interventions.

In this work we investigate a variation of the manual initialization approach. We use the Microsoft Kinect system[4] to interactively explore the six degree-of-freedom parameter space. The new method adds a Kinect-based interface to our 2D/3D initialization tool, replacing the use of mouse-based operations.

2. METHODS

We have developed a graphical user interface which allows the user to interactively position a 3D volume in space, with the intent of aligning it to an X-ray image. When multiple X-ray images are available the user manipulates the volume pose until it is aligned with all of the images. Figure 1 shows our graphical user interface which integrates both mouse-based and gesture-based interaction for 2D/3D registration.

The user starts by loading a 3D data set, CT or MR, and a set of X-ray images. Initialization is then performed in two steps, automatic positioning limited to translation, and interactive pose determination.

Automatic positioning is realized by aligning the geometrical center of the 3D data with the intersection of the principal rays of all X-ray cameras. In practice, the X-ray principal 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. It should be noted that automatic positioning only accounts for translation.

The user then selects an X-ray image as a reference view and interactively performs six degree-of-freedom spatial manipulation, rotations and translations, on the 3D data so that its 2D rendering becomes visually more similar to the X-ray image. The 2D renderings are generated in real time from the 3D data using the graphics processing unit to provide interactive performance. Interaction usually starts by grossly positioning the 3D data so that it is aligned with all X-rays using only translations; then, the alignment is iteratively refined by applying rotations and translations in each of the X-ray views.

To provide intuitive feedback on screen during the process of initialization, a user interaction, rotation or translation, yields a *virtual* pose update to the selected X-ray view instead of the 3D data. The *actual* pose

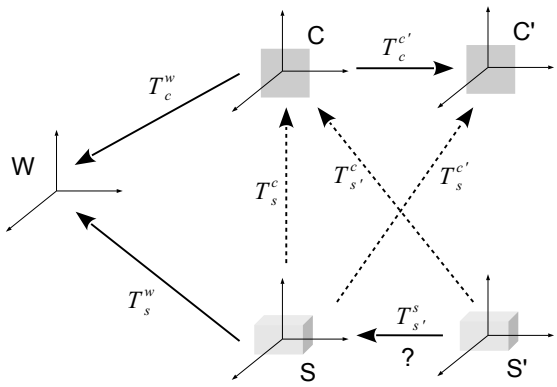

Figure 2. Calculation of pose update to the 3D data for a user interaction.

update to the 3D data needs to be calculated. Let W be the world, C and C' be the X-ray view at its current and virtual new positions, S and S' be the 3D data at its current and actual new positions, $T_c^{c'}$ be the virtual pose update to the X-ray view due to user interaction, and $T_{s'}^{s}$ be the actual pose update to the 3D data to be calculated. Figure 2 illustrates the relationship among those entities. Based on the fact that $T_{s'}^{c} = T_s^{c'}$, as well as $T_{s'}^{c} = T_s^c T_{s'}^s$ and $T_s^{c'} = T_c^{c'} T_s^c$, the pose update to the 3D data can be calculated as

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

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

where T_s^c represents the current pose of the 3D data in the coordinate frame of the X-ray view, T_c^w is the known pose of the X-ray view in the world coordinate frame and T_s^w is the current pose of the 3D data in the world coordinate frame. The new pose of the 3D data is then obtained as $T_{s'}^w = T_s^w T_{s'}^s$.

2.1 Mouse-based interaction

In the conventional initialization method, the operator uses a mouse to interact with controls on the graphical user interface, which include an X-ray view selector, three wheels and dial for the rotation, and three wheels and dial for the translation, as shown in Figure 1. In this interaction mode, the user can easily switch among or operate different controls: clicking on an entry to select an X-ray view, or dragging a wheel or dial to translate along or rotate about one of the three X-ray view coordinate axes.

2.2 Gesture-based interaction

In this interaction mode, we utilize the Microsoft Kinect.[5] Based on user motion and gestures we simulate the mouse-based operations described above. To provide reliable gesture and motion detection, the sensor data is

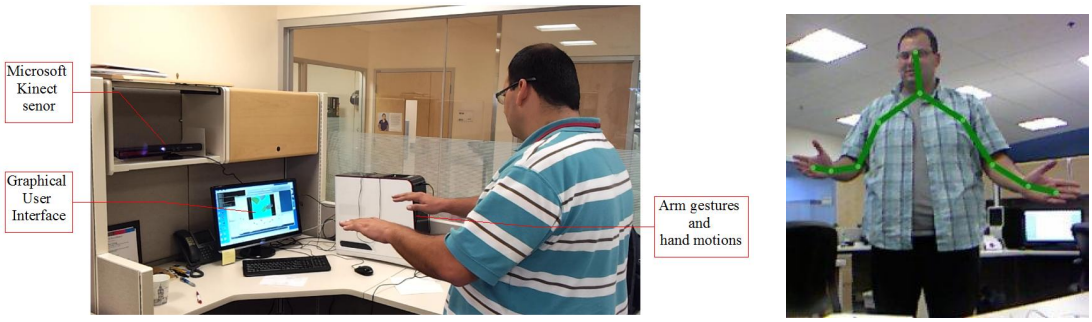

Figure 3. Gesture based interaction with the Microsoft Kinect, and the corresponding skeletal joint locations, used to define gestures, overlaid onto the video.

filtered and processed. This includes smoothing to remove noise and scaling to provide fine control of motion on screen. The Kinect automatically identifies the closest most active user and reports the joint positions of the associated skeletal model, as shown in Figure 3. We use the Kinect in "seated mode", which means that only the joint positions above the hip are reported. This is more appropriate for use in the OR where the surgeon's legs may not be visible if they are standing behind the operating table.

Gestures are based on the reported joint positions. We define the following gestures for interaction:

1. Right-arm above the shoulder moving from right to left: select the next X-ray as the reference.
2. Right-arm above the shoulder moving from left to right: switch the manipulation mode between translation and rotation.
3. Left-arm extended forward: enable motion tracking of the right hand.
4. Left-arm in straight down position and right-arm fast moving from right to left: select the next coordinate axis as the rotation axis in the rotation mode.

The translation and rotation are obtained by tracking the motion of the right hand. For translation, all three components of the hand motion are used. This differs from the mouse-based interaction, which performs translation with one degree-of-freedom at a time. For rotation, only the magnitude of the motion is used with regard to the active axis, disabling compound rotations.

3. EXPERIMENTS

As our interaction approach is based on gestures designed for a specific task we first needed to assess the amount of training it takes to confidently interact with the program. We then evaluate the use of our program for interactive initialization, using accuracy and interaction time as the quality measures.

3.1 User training for gesture-based initialization

The efficiency of using our Kinect-based initialization tool depends on familiarity with the Kinect device and the gestures we have defined.

To assess the amount of training required to master this mode of interaction we conducted a training session. We created a simulated reference data set from a diagnostic CT of a distal radius osteotomy. We simulated X-ray images from known camera poses. The X-ray images are easily interpreted, and the pose of the 3D data is easily discerned. Figure 4 shows the overlays between the simulated X-ray images and an interactive volume rendering of the 3D data at an arbitrary pose.

Three participants took part in the training session. One participant was the developer of the gesture based interaction functionality and was familiar with the Kinect, one had experience in using Kinect with computer

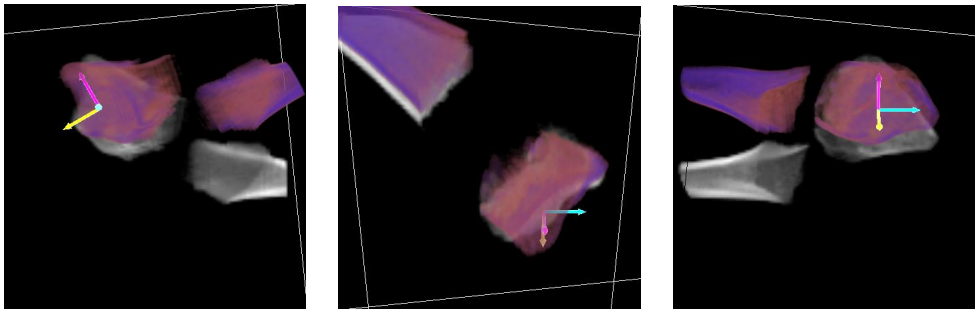

Figure 4. Data set used for user training for gesture-based initialization. Three X-ray images from orthogonal views were simulated from the CT of segmented fracture fragments. The overlaid volume renderings were generated from the 3D data.

games but was new to our set of gestures, and one had no previous experience with Kinect. During the training, random poses of the 3D data were generated, and each participant performed gesture-based initialization to re-position the 3D data until they were certain that the volume was positioned correctly based on visual inspection. We recorded the elapsed time for ten trials for each participant, and the results were analyzed.

3.2 Initialization Evaluation

3.2.1 Data

We evaluate our initialization approach using three publicly available reference data sets for 2D/3D registration. The first data set[6] is from the Image Science Institute (ISI), Netherlands, and consists of images from a spine phantom containing three vertebra. The second data set[7] is from the University of Ljubljana, and consists of five lumbar vertebra. The third data set[8] 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.

For each of the data sets, we selected two X-ray images, one CT, one MR, and the reference transformations for the CT and MR. The reference transformations position the 3D images in the world coordinate frame to match the corresponding X-ray images. Figure 5 shows all X-ray images used in this study, as well as the corresponding volume renderings of CT and MR in the reference poses.

3.2.2 Evaluation design

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

For each testing case, we performed both mouse-based and gesture-based initialization, and recorded the initial error, final error and elapsed time.

Two participants took part in this evaluation: one was the developer of our gesture-based interaction approach and the other had no previous experience using the Kinect device.

To accurately position the 3D data we assume that the user is familiar with the anatomical structure. This assumption is valid for clinicians but was not valid for the participants in this experiment. To address this issue we allowed each participant to spend ten minutes studying the anatomy on screen before starting the experiments. During this period, the participant arbitrarily manipulated the 3D data in order to obtain an understanding of the spatial structure of the anatomy.

3.2.3 Accuracy evaluation

Initialization errors were evaluated using a set of targets within the 3D data (CT or MR). To obtain the targets, the 3D data was interactively thresholded such that points belonging to the bone surface were selected. 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, which is defined as

$$mTRE(r;S) = \frac{1}{N}\sum_{i=1}^{N} \parallel T_r p_i - T_g p_i \parallel, \tag{3}$$

where p_i is a target at index i, and N is the total number of targets.

Errors before and after interactive registration were recorded. Note that the initial errors are the errors after automatic positioning of the volume (see Section 2) and not those obtained from perturbation.

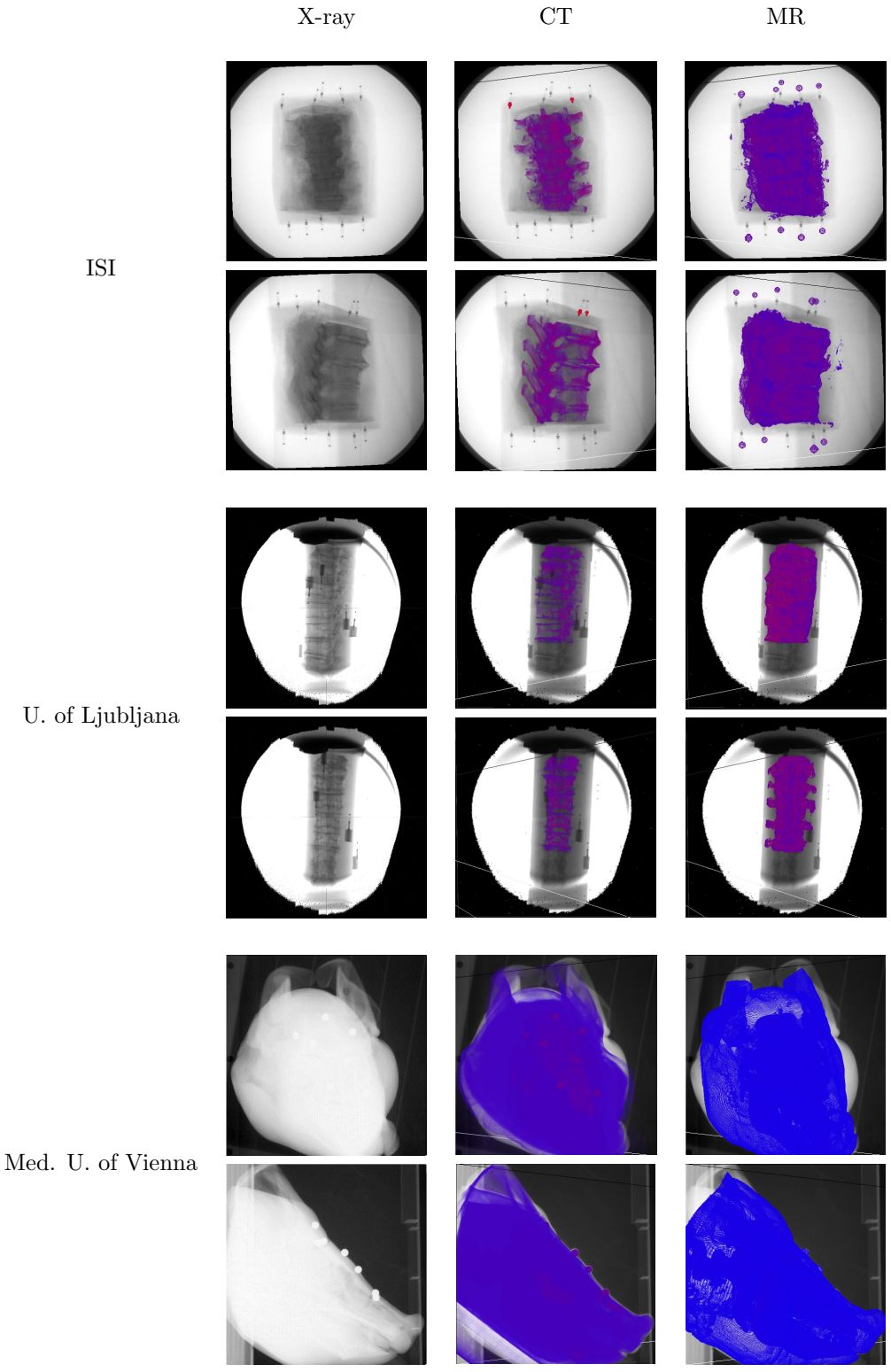

Figure 5. Testing data used in this study. The X-ray images for the ISI and Vienna data sets were provided in the anterior-posterior and lateral views. The Ljubljana data set provided 18 evenly spaced X-ray images around the spinal cord, and we selected two images (000 and 006) in our study. Original 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.

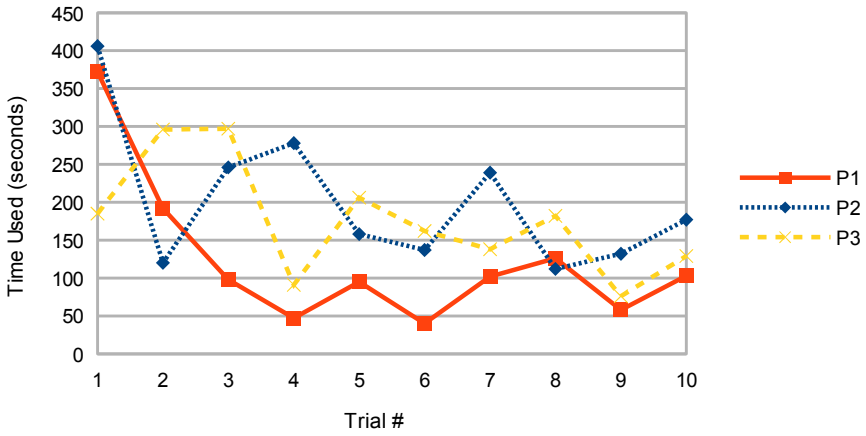

Figure 6. Distribution of used time for the user training study. P1: the developer. P2: the participant without Kinect experience. P3: the participant with Kinect experience.

4. RESULTS AND DISCUSSION

4.1 User training for gesture-based initialization

Figure 6 summarizes the results from our training session. From these results we have two findings: First, initially a single Kinect-based session took 5-6 minutes for all participants. After ten trials, it was reduced to below 3 minutes. In total ten trials took 21, 29 and 33 minutes for the three participants. We thus conclude that, with about half an hour of training, a new user should be able to confidently interact with our gesture-based initialization program. Second, for users who have previous experience with the Kinect, such as P1 and P3, it takes a shorter time to master the use of gestures to interact with our program.

4.2 Initialization for X-ray/CT registration

Figure 7 summarizes the experimental results for X-ray/CT initialization. We observe the following:

- Participant 1 versus 2: from the column mTRE, there is no obvious evidence showing that one participant consistently performed better than the other for all data sets, and all experiments completed with a final mTRE less than 10mm. Also, the mTRE variances for the two participants were comparable except for the case of gesture-based experiments for the ISI data set. In that particular case, participant 1 (the developer) had a good understanding of the anatomy after mouse-based experiments, so he could finish all Kinect-based trials with a relatively constant time. When looking at the column time, participant 1 used less time to complete the experiments, partly because he allowed larger initialization errors (see the column mTRE), partly because he, as the algorithm developer, was more familiar with the initialization tool. The above observations showed that all users could achieve equally good initialization results with our tool, and a user's experience with Kinect, understanding about the anatomy, and so on, reflect in the initialization error and time.
- Mouse-based versus gesture-based: from the column mTRE, except for the Vienna data set, there is no obvious evidence showing that one method is superior to the other in terms of final mTRE. In the case of the animal data, the gesture-based errors were noticeably larger than the mouse ones. This was because in this data set the mTRE was sensitive to small rotation errors. Using the mouse-based method allowed for better fine-tuning support to handle small pose errors. In general, the two methods achieve comparable initialization accuracy. When looking at the time, the gesture-based experiments took slightly longer time. We believe that this discrepancy can be reduced when the users gain more experience with the new method.

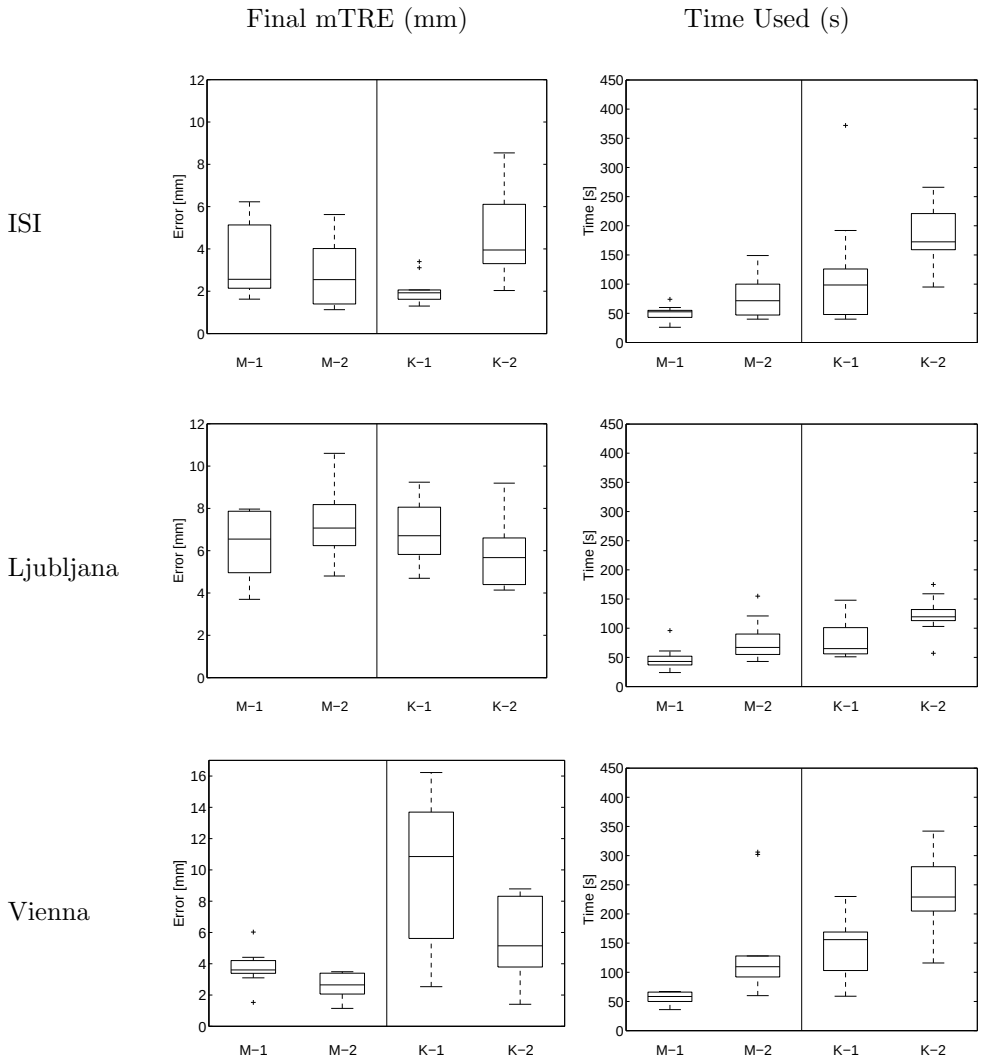

Figure 7. Experimental results for X-ray/CT initialization. M-1 and M-2: mouse-based experiments for participants 1 and 2. K-1 and K-2: Kinect-based experiments for participants 1 and 2. Participant 1 was the algorithm developer.

- For different type of data sets: from the column mTRE, we see that the shape of the anatomy had a direct impact on final errors. This is more obvious for the ISI and Ljubljana data sets - both were for a spinal phantom. When looking at the column time, we also see that the shape of anatomy had a impact on the time, though the impact was not significant.

4.3 Initialization for X-ray/MR registration

Fig. 8 summarizes the experimental results for X-ray/MR initialization. Compared to X-ray/CT initialization, the difference between X-ray images and 2D renderings of the 3D data, i.e. MR, is much larger, so we have different observations:

- Participant 1 versus 2: we can see that participant 2 performed slightly better for the ISI and Ljubljana data sets, in terms of both final mTRE and mTRE variance. This is mainly because he spent longer time (see column time) in order to achieve better initialization. For the Vienna data set, participant 1 (the developer) used significantly shorter and more constant time for the mouse-based experiments, mainly because he allowed higher initialization errors (see column mTRE, M-1 and M-2 in the Vienna plot). In

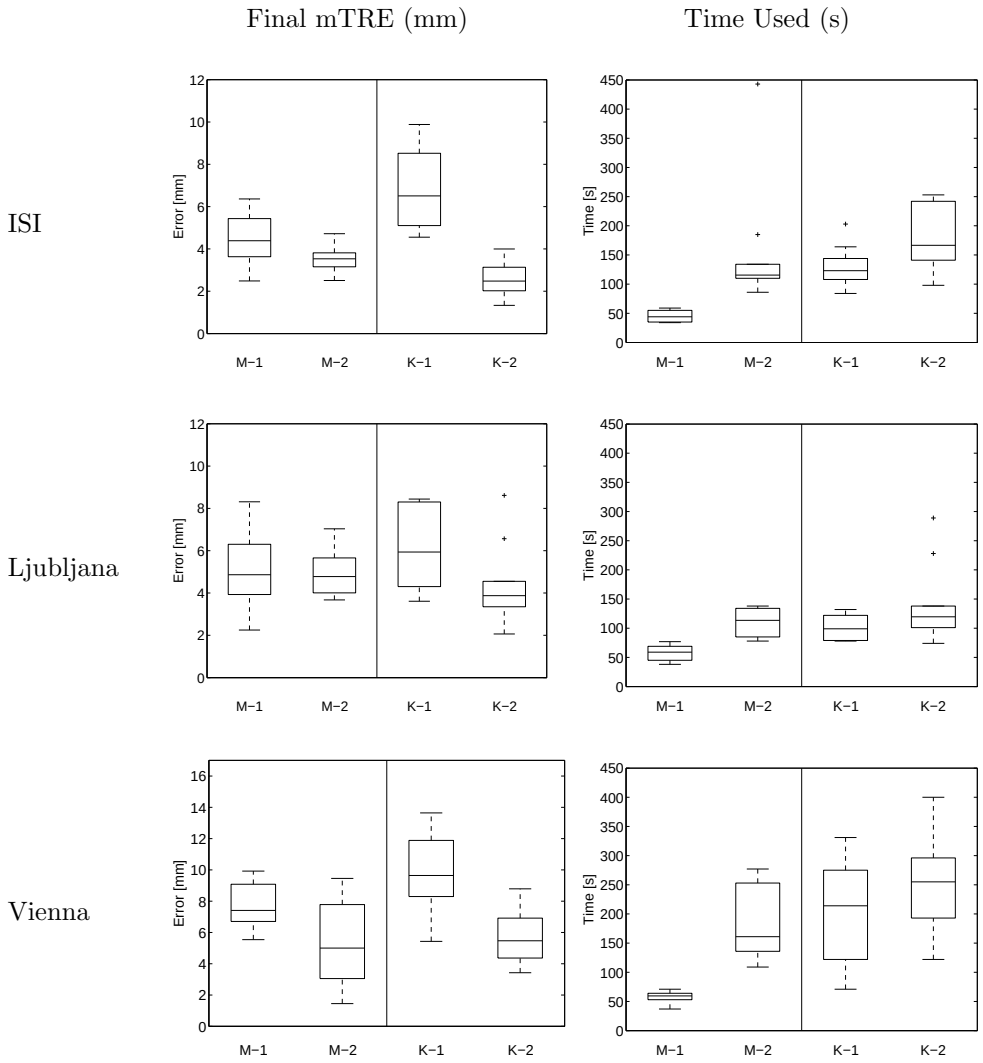

Figure 8. Experimental results for X-ray/MR initialization. M-1 and M-2: mouse-based experiments for participants 1 and 2. K-1 and K-2: Kinect-based experiments for participants 1 and 2. Participant 1 was the algorithm developer.

general, we can still say that all users could achieve equally good initialization results with our tool, and user's experience with our tool and the data set affects the initialization error and time.

- Mouse-based versus Kinect-based: we have similar observations as the X-ray/CT initialization, that is, no one method was superior to the other in terms of final mTRE, and Kinect-based method took slightly longer time because it uses a new interaction mode.
- For different type of data sets: not like the X-ray/CT initialization, there was no evidence showing that the errors for the Ljubljana data set were larger than the ISI ones due to more complicated anatomical structures. This discrepancy was mainly because the volume rendering of MR does not closely simulate X-ray images, and it is difficult to find reliable structures that are visible on both types of images to guide the initialization. One thing was obvious, that is, the final errors for X-ray/MR initialization were larger than the ones for X-ray/CT initialization (see Fig. 7). This is also due to the large difference between X-ray images and 2D renderings of MR.

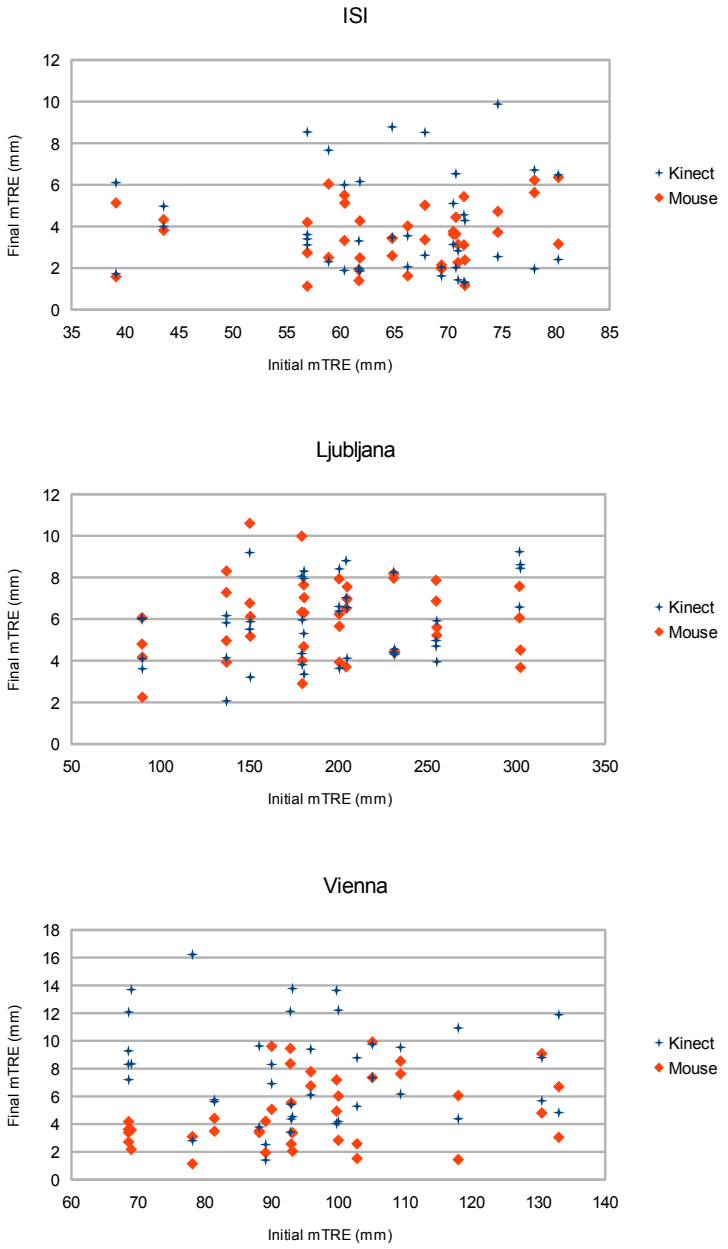

Figure 9. Initial versus final initialization errors (results for X-ray/CT and X-ray/MR initialization and from participants 1 and 2 were combined).

Table 1. Summary of results (results from participants 1 and 2 were combined, numbers read as mean ± st.dev.).

	Initial mTRE (mm)	Mouse-based		Kinect-based	
		Final mTRE (mm)	Time (s)	Final mTRE (mm)	Time (s)
ISI	64.8 ± 10.2	3.6 ± 1.5	82 ± 69	4.0 ± 2.4	153 ± 69
Ljubljana	193.1 ± 58.5	6.0 ± 1.9	73 ± 34	5.9 ± 1.9	112 ± 47
Vienna	95.3 ± 18.0	4.9 ± 2.5	110 ± 76	7.7 ± 3.6	210 ± 80

4.4 Summary

Figure 9 shows the relationship between initial and final errors for individual data sets. We can see that, for all randomly generated initial poses, our gesture-based method was able to provide initialization successfully, and the final errors for mouse-based and gesture-based methods were in the same order.

Table 1 lists the statistics about the initialization errors and time for all experiments performed within this work. By further consolidating the errors and time for all data sets, we have following observation: with initial target registration errors of 117.7 ± 28.9 mm, a user was able to achieve final errors of 5.9 ± 2.6 mm within 158 ± 65 sec using the gesture-based approach, compared to 4.8 ± 2.0 mm and 88 ± 60 sec for the mouse-based approach. These alignment errors are sufficiently accurate to initiate most available X-ray/CT registration algorithms, and considered to be good for X-ray/MR registration too.

5. CONCLUSION

This paper described an interactive initialization approach for 2D/3D intra-operative rigid registration. We use the Microsoft Kinect to interactively position the 3D data, MR or CT, so that rendered views correspond to the intra-operatively acquired X-ray images. This enables us to initialize a 2D/3D registration algorithm without requiring the use of fiducials and allows the physician to control the process while adhering to the requirements of a sterile operating room environment.

Experimental results showed that, initialization using gesture-based interaction results in similar accuracy to mouse-based interaction. For users that have no prior experience using the Kinect, half an hour of training was sufficient to master the gesture-based interaction method, though the interaction time was slightly longer than that of an experienced user. Our experiments also showed that the initialization error and time depend both on the user's experience with our tool and the shape of the anatomical structure, or the user's perception of the anatomy. We believe that, after the user gains more experience with our tool, the initialization error and time can be further reduced.

ACKNOWLEDGMENTS

We would like to thank P. Cheng for his help in the evaluation studies.

REFERENCES

[1] Markelj, P., Tomaževič, D., Likar, B., and Pernuš, F., "A review of 3D/2D registration methods for image-guided interventions," *Medical Image Analysis* **16**(3), 642–661 (2012).

[2] van der Bom, M. J., Bartels, L. W., Gounis, M. J., Homan, R., Timmer, J., Viergever, M. A., and Pluim, J. P. W., "Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation," *Med. Phys.* **37**(4), 1884–1892 (2010).

[3] Yaniv, Z., "Rigid registration," in [*Image-Guided Interventions: Technology and Applications*], Peters, T. and Cleary, K., eds., ch. 6, Springer-Verlag (May 2008).

[4] Jana, A., [*Kinect for Windows SDK Programming Guide*], Packt Publishing (2012).

[5] Lejeune, A., Van Droogenbroeck, M., and Verly, J., "The secrets of the Kinect ... in depth!." Presentation at the Conference of Professional Forum (3D Stereo Media), Liège, Belgium (December 2011).

[6] van de Kraats, E. B., Penney, G. P., Tomaževič, D., van Walsum, T., and Niessen, W. J., "Standardized evaluation methodology for 2-D3-D registration," *IEEE Trans. Med. Imag.* **24**(9), 1177–1189 (2005).

[7] Tomaževič, D., Likar, B., and Pernuš, F., "Gold standard data for evaluation and comparison of 3D/2D registration methods," *Computer Aided Surgery* **9**(4), 137–144 (2004).

[8] Pawiro, S., Markelj, P., Pernus, F., Gendrin, C., Figl, M., Weber, C., Kainberger, F., Nbauer-Huhmann, I., Bergmeister, H., Stock, M., Georg, D., Bergmann, H., and Birkfellner, W., "Validation for 2D/3D registration I: A new gold standard data set.," *Med. Phys.* **38**(3), 1481–90 (2011).