Pose-aware C-arm for automatic re-initialization of interventional 2D/3D image registration

Javad Fotouhi[1], **Bernhard Fuerst**[1], **Alex Johnson**[2], **Sing Chun Lee**[1], **Russell Taylor**[3], **Greg Osgood**[2], **Nassir Navab**[1,4], and **Mehran Armand**[2,5,6]

[1]Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD, USA

[2]Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, MD, USA

[3]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

[4]Computer Aided Medical Procedures, Technische Universität München, Munich, Germany

[5]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA

[6]Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

Abstract

Purpose—In minimally invasive interventions assisted by C-arm imaging, there is a demand to fuse the intra-interventional 2D C-arm image with pre-interventional 3D patient data to enable surgical guidance. The commonly used intensity-based 2D/3D registration has a limited capture range and is sensitive to initialization. We propose to utilize an opto/X-ray C-arm system which allows to maintain the registration during intervention by automating the re-initialization for the 2D/3D image registration. Consequently, the surgical workflow is not disrupted and the interaction time for manual initialization is eliminated.

Methods—We utilize two distinct vision-based tracking techniques to estimate the relative poses between different C-arm arrangements: (1) global tracking using fused depth information and (2) RGBD SLAM system for surgical scene tracking. A highly accurate multi-view calibration between RGBD and C-arm imaging devices is achieved using a custom-made multimodal calibration target.

Results—Several in vitro studies are conducted on pelvic-femur phantom that is encased in gelatin and covered with drapes to simulate a clinically realistic scenario. The mean target registration errors (mTRE) for re-initialization using depth-only and RGB+depth are 13.23mm and 11.81mm, respectively. 2D/3D registration yielded 75% success rate using this automatic re-initialization, compared to a random initialization which yielded only 23% successful registration.

Conclusion—The pose-aware C-arm contributes to the 2D/3D registration process by globally re-initializing the relationship of C-arm image and pre-interventional CT data. This system performs inside-out tracking, is self-contained, and does not require any external tracking devices.

Keywords

2D/3D registration; C-arm; Initialization; Intra-intervention; RGBD camera

Introduction

C-arm 2D fluoroscopy is the crucial imaging modality for several image-guided interventions. During these procedures, the C-arm is frequently re-positioned to acquire images from various perspectives of a target anatomy. Performing surgery solely under 2D fluoroscopy guidance is a challenging task as a single point of view lacks the information required to navigate complex 3D structures, making the intervention and procedure extremely difficult.

Acquiring cone-beam computed tomographic (CBCT) volumes which provides 3D information of the anatomy is a solution to this problem [7,9,19]. However, CBCT acquisition does not yield real-time feedback. A considerable number of scans would need to be performed at every step of the procedure, requiring the surgical site to be prepared for each scan [5].

Pre-interventional 3D patient data can be fused to the C-arm images and augment the intra-interventional fluoroscopy images with 3D information. This multimodal image fusion is used in several different interventions, particularly in radiation therapy [11], cardiac and endovascular procedures [21], and orthopedics and trauma interventions [26]. Additionally, other image-guided orthopedic procedures such as femoroplasty [25], (robot-assisted) bone augmentation, K-wire and screw placement in pelvic fractures, cup placement for hips, and surgeries for patients with osteonecrosis can benefit from the registration of pre-interventional data with real-time intra-interventional C-arm images. Therefore, there is a need for robust registration techniques that enable a seamless alignment between 3D and 2D data. We present a vision-based method to automatically re-initialize the registration of pre- and intra-interventional data using the RGBD camera attached to the C-arm.

Background

Intensity-based image registration is one group of image-based registration techniques and of special interest in this paper. Registration is typically performed by simulating 2D radiographs, so-called digitally reconstructed radiographs (DRRs), from pre-interventional patient data, and matching them with the intra-interventional C-arm image [13,17,20]. Intensity-based registration becomes challenging where bony structures in the pre- and intra-interventional data differ due to deformations caused by the surgery. C-arm image registration is in particular a complex problem due to the limited views of the C-arm images. Atlases of the target anatomy and deformable shape statistics are suggested to support the registration in such cases of image truncation [29]. In spite of promising results from using shape models, in several trauma and orthopedic cases where the bony anatomy is severely

damaged and deformed, the registration is yet prone to failure. Last but not least, intensity-based registration is very sensitive to initialization due to the limited capture range of the intensity-based similarity cost. These are the challenges that prohibit the 2D/3D registration to become a part of standard surgical routine.

Related works

Reliable registration is often performed using external navigation systems [4]. In navigation-guided fluoroscopy, it is a common practice to drill the fiducials into the bone to maintain the registration in the presence of patient movement [24]. Generally, these fiducials are implanted before the 3D pre-interventional data are collected, and will remain inside the anatomy until after the intervention. Fiducial implantation requires a separate invasive surgery and increases the risk of fracture in the osteoporotic bones.

Fiducial-based C-arm tracking was used for intensity-based image registration in [24]. This method achieves submillimeter accuracy and, however, requires the implantation of a custom-made in-image fiducial. The mean target registration error (mTRE) is 0.34mm for the plastic bone phantom with 90° rotation, and 0.99mm for the cadaveric specimen with images 58.5° apart. In [16], anatomical features or beads on patient surface are located in the C-arm images and used for C-arm pose estimation. This work tackles the problem of intra-interventional calibration and reconstruction and does not address registration to pre-interventional data. Despite the high registration accuracy of navigation-guided systems near the fiducial markers, the registration error and uncertainty increase in distant areas. Furthermore, these systems have a complex setup, occupy additional space, and change the surgical workflow. Last but not least, the line of sight issues limit the free space in the surgical site.

In order to initialize the 2D/3D registration for several number of C-arm images, Uneri et al. [30] proposed to use each successful registration of a C-arm image to 3D data as the initialization for the latterly C-arm image. This work avoids the use of any external trackers by solely relying on image-based registration. But on the other hand, in order to robustly initialize the registration, the allowed displacement between consecutive C-arm images was limited to only 10mm.

The initial alignment and outcome verification are introduced as two main bottlenecks of the registration problem [31]. In order to address the problem associated with initialization of a 2D/3D registration task, Gong et al. [14] suggested an interactive initialization technique where the user performs the alignment by utilizing a gesture-based interface or an augmented reality environment together with a navigation system. The tedious initialization procedure makes the system impractical in a surgical setting.

Generalized Hough Transform is used in [32] to learn large number of 2D templates over a variety of 3D poses. During intervention, this information is used to estimate the 3D pose from the 2D C-arm images and initialize the 2D data with respect to 3D pre-interventional CT. Projection-slice theorem and phase correlation are also used to estimate the initialization of the 2D/3D registration problem [2].

Proposed solution for automatic re-initialization of registration

In this work, we use vision information from an RGB-depth (RGBD) camera mounted near the detector of a mobile C-arm to estimate the pose of the C-arm relative to the patient. We refer to this C-arm as *pose-aware*. The proposed C-arm tracking will result in estimating the projection geometry of the C-arm relative to the surgical scene at arbitrary poses of the C-arm. The tracking information will be used to transform the 2D C-arm images globally near their correct alignment with respect to the pre-interventional data. Next, intensity-based registration is utilized to align images locally. In the spirit of using one modality (ultrasound) to initialize the registration of another imaging modality (endoscopic video images), the method suggested in [34] resembles our work. Our methodology aims at improving the automation during C-arm guided interventions. This technique is more effective when the registration has to be repeated for multiple C-arm poses. The workflow (Fig. 1) initiates by performing an initial registration of the C-arm to the pre-interventional data. The registration loop is then closed by using this pose as the reference for the tracking of different C-arm arrangements. This system is self-contained and requires a single factory calibration, and the surgical workflow remains intact.

Materials and methods

The proposed pose-aware C-arm uses a co-calibrated RGBD camera to track the surgical site in the course of the C-arm re-positioning. The joint calibration of C-arm and RGBD imaging devices is discussed in "Calibration of RGBD camera and C-arm" section. We then introduce two distinct tracking methods based on depth-only and RGB+depth data for pose estimation of the C-arm ("Estimation of C-arm trajectory using visual tracking" section). "Pose-aware C-arm for intensity-based 2D/3D image registration of pre- and intra-interventional data" section presents the intensity-based registration of 2D intra-interventional C-arm images and 3D pre-interventional CT using DRRs and the automatic re-initialization using the pose-aware C-arm system.

Calibration of RGBD camera and C-arm

The RGBD camera is attached rigidly to the C-arm, allowing tracking performed using the camera to be translated to the pose of the X-ray source and detector. To obtain the relationships between RGB camera, depth sensor, and X-ray source, we perform a standard multi-view calibration [10,35]. To enable this technique, we deploy a multimodal checkerboard, which can be observed by the RGB camera, the infrared camera in the depth sensor, and the X-ray imaging device [8]. Note that the X-ray image is usually flipped adjust of the X-ray source to be below the table, while the clinician is observing the patient from the top. Finally, we obtain the transformation $^X\mathbf{T}_{RGBD}$ from RGBD sensor, where we define the depth component to be the origin, to the X-ray source.

Estimation of C-arm trajectory using visual tracking

The tracking of C-arm source during C-arm re-positioning is performed by first observing the surgical site (patient) using the RGBD camera. These relationships, denoted as $^P\mathbf{T}_{RGBD}$

and $^P\mathbf{T}_{RGBD'}$, are shown in Fig. 2. The relative transformation between the RGBD camera poses is then derived as:

$$^{RGBD'}\mathbf{T}_{RGBD} =^P \mathbf{T}_{RGBD'}^{-1} \cdot {}^P \mathbf{T}_{RGBD} \quad (1)$$

Since the RGBD camera is rigidly mounted on the C-arm, the multi-view relation between the RGBD origin and X-ray source remains fixed ($^X\mathbf{T}_{RGBD} =^{X'} \mathbf{T}_{RGBD'}$). We can then obtain the relative displacement of the X-ray source as:

$$^{X'} \mathbf{T}_X =^{X'} \mathbf{T}_{RGBD'} \cdot {}^{RGBD'} \mathbf{T}_{RGBD} \cdot {}^X \mathbf{T}_{RGBD}^{-1} \quad (2)$$

In the following, we discuss two distinct vision-based methods for tracking the surgical site with depth-only and RGB+depth data.

Tracking and surface reconstruction using depth data

We deploy an algorithm to automatically estimate the depth sensor trajectory and create a smooth dense 3D surface model of the objects in the scene [23]. The relative camera poses are acquired by iteratively registering the current depth frame to a global model. This iterative stage is performed using multi-scale iterative closest point (ICP), where depth features among frames and the global model are matched using projective data association [28].

RGBD-based simultaneous localization and mapping

Modern RGBD cameras are small devices which provide color information as well as depth data. Therefore, greater amount of information is extracted and processed compared to a single RGB or infrared camera. We aim at using depth and color information concurrently and acquire 3D color information from the surgical site. Ultimately, we use this knowledge to track the C-arm relative to the patient. For this purpose, we use an RGBD SLAM system introduced in [3]. The underlying tracking method uses visual features (e.g., SURF [1]) extracted from the color images and constructs feature descriptors. The feature descriptors are matched among the consecutive frames to form a feature correspondence list. The initial transformation is then computed using the correspondence list and a random sample consensus (RANSAC) method [6] for transformation estimation. The outcome is later used to initialize ICP to further refine the mapping between consecutive frames. Lastly, to compute a globally consistent trajectory, a pose graph solver is used that optimizes the trajectory using a nonlinear energy function [15,18].

Pose-aware C-arm for intensity-based 2D/3D image registration of pre- and intra-interventional data

Intensity-based 2D/3D image registration is an iterative method which 6 DOF parameters are estimated for a rigid body transformation that brings the input 2D and 3D data to a

common coordinate frame. The key component of this process is the creation of many 2D artificial radiographs, so-called digitally reconstructed radiographs (DRRs), from the CT data and comparing them against the C-arm X-ray image. This process continues until the similarity cost between the DRR and X-ray images is maximized using an iterative optimization.

CT image is a collection of voxels representing X-ray attenuation. In order to generate a DRR from a given CT data, 6 DOF rotation and translation parameters are used together with the intrinsics of the C-arm to position a virtual source and a DRR image plane near the CT volume [33]. The intrinsic calibration of the C-arm allows us to estimate the source to detector arrangement and use this information to generate DRR images [22]. Next, by means of raycasting [12], the accumulation of intensities along each ray (that originates from the virtual source and intersects with the image plane) is computed. The accumulated values then form a DRR image. In this work, we first perform a single image-based C-arm image to CT registration using manual initialization. There-after, we track the C-arm new pose relative to the surgical site. The tracking results of the C-arm are later used as the 6 DOF initialization parameters of the new coming C-arm images.

Next, the C-arm X-ray images are converted to linear accumulation of X-ray attenuation by means of log-normalization. Hence, the DRRs and detector images become linearly proportional. Furthermore, we apply a circular mask on the detector images to discard the saturated areas near the detector boundaries and only include relevant information in the similarity cost.

After estimating the re-initialization transformation $\mathbf{T}_0$ obtained as results of the composition of transformations $^{X'}\mathbf{T}_X$ estimated by the pose-aware system, and that of the first pose of the C-arm, the DRR image is generated and compared against the C-arm X-ray image I_X using a similarity cost S. In this work we utilize the normalized cross-correlation (NCC) as the similarity cost defined as:

$$\text{NCC} := S((\mathbf{R}, \mathbf{t}) | \mathbf{T}_0, I_X, I_D) = \sum^{\left|\Omega_{X,D}\right|} \frac{I_X \cdot I_D^{((\mathbf{R}, \mathbf{t}) | \mathbf{T}_0)}}{\sigma_X \sigma_D} \quad (3)$$

where I_X and I_D are the mean-normalized X-ray and DRR images, $(\mathbf{R}, \mathbf{t})$ are the rotation and translation parameters used to form the DRR image, $\Omega_{X,D}$ is the common spatial domain of the two images, and σ_X and σ_D are the standard deviations of the X-ray and DRR image intensities in $\Omega_{X,D}$.

The final parameters $(\widehat{\mathbf{R}}, \widehat{\mathbf{t}})$ are estimated by an iterative optimization:

$$[\widehat{\mathbf{R}}, \widehat{\mathbf{t}}] = \underset{\mathbf{R}, \mathbf{t}}{\arg \max}\, S((\mathbf{R}, \mathbf{t}) | \mathbf{T}_0, I_X, I_D) \quad (4)$$

The optimization is performed using bound constrained by quadratic approximation (BOBYQA) algorithm [27].

Experimental validation and results

Experimental setup

X-ray images are acquired using an Arcadis Orbic 3D mobile C-arm from Siemens Healthineers with the detector size of 230mm×230mm. In order to remove low noise, the X-ray images are captured as digital radiographs (DRs) where a weighted average filter is applied to the images. The DR images are acquired using 0.2–23mA and 40–110kV. For safety purposes, the maximum power for taking a DR image is set to 1000W. The CT scan data are acquired using a Toshiba Aquilion One CT scanner, where the slice spacing and thickness are 0.5mm, and the volume elements are created using 16bits. An Intel RealSense SR300 camera from Intel Corporation is rigidly mounted on the image intensifier using a custom-made 3D printed mount. The SR300 has a small form factor ($X = 110.0 \pm 0.2$ mm, $Y = 12.6 \pm 0.1$mm, $Z = 3.8$–4.1mm), and integrates a full-HD RGB camera, an infrared projector, and an infrared camera. To ensure a wide visibility range of the patient surface, the RGBD camera is positioned with a particular tilt angle that the center of the image is nearly aligned with the C-arm iso-center. The C-arm imaging device is connected via Ethernet, and the RGBD camera is connected via powered USB 3.0 to the development PC which runs the tracking and the registration software. The tracking is performed in real-time, and the 2D/3D registration module is fully parallelized using graphical processing unit acceleration.

Results

In the following, we evaluate the pose-aware C-arm system by presenting the calibration accuracy, tracking results, and 2D/3D registration performance with and without the automatic re-initialization.

Calibration outcome

A 5×6 checkerboard calibration target is employed to calibrate the X-ray, RGB, and infrared (depth channel of the RGBD camera) imaging devices. The length of each side of a single checkerboard square is 12.66mm, and the total dimensions of the checkerboard are 63.26 and 75.93mm. The checkerboard size is carefully selected according to the field of view of the C-arm and the RGBD camera images to ensure full visibility among a variety of checkerboard poses. Next, 72 X-ray/RGB/IR images were simultaneously recorded to perform the calibration. A subset of images that produced high reprojection error were discarded as outliers. Ultimately, we used 42 image pairs and performed stereo calibration between RGB camera and X-ray source where the stereo mean reprojection error was 0.86 pixels. The stereo calibration in this work does not refer to the calibration of a single moving camera, instead it refers to estimation of the extrinsic parameters of two stationary cameras (RGB and X-ray imaging devices). RGB and X-ray images had individual mean reprojection errors of 0.75 and 0.97 pixels, respectively.

RGB and IR cameras were calibrated using 59 images, and the mean stereo reprojection error was 0.17 pixels, with individual mean reprojection errors of 0.18 and 0.16 pixels for each camera, respectively.

Tracking accuracy

The accuracy of the vision-based tracking methods is evaluated by rotating the C-arm around different axis, computing the pose of the X-ray source, and comparing the outcome to the ground truth provided by an external optical tracking system. An exemplary rotation is shown in Fig. 3, where the C-arm orbits around the surgical bed.

The optical position sensor tracks a single-faced passive rigid body fiducial mounted on the C-arm which is pre-calibrated to the X-ray source. The tracker is a Polaris Vicra System from Northern Digital Inc., where the measurement area for the minimum distance (557 mm away from the tracker) is 491 mm × 392 mm, and 938 mm × 887 mm for the maximum distance (1336 mm) from the tracker.

The C-arm is rotated along three main axis, and the tracking results from the depth-only algorithm and RGBD SLAM are recorded and compared to an optical tracking system. First, C-arm is orbited 45° around the cranial/caudal axis. The tracking errors are shown in Fig. 4a, b in the form of Euclidean distance, as well as the rotational error. The rotational error is computed as the norm of mal-rotation angle along the three axis with respect to the ground truth. Next, the C-arm is swiveled by ±10°, and the tracking errors are plotted in Fig. 4c, d. Lastly, the C-arm is rotated to anterior-posterior oblique view with ±30°. The results of this experiment are shown in Fig. 4e, f.

Note that the C-arm rotation does not only result in the rotation of the X-ray source, but also by construction of the device a significant translation is applied to the X-ray source. In an exemplary case, considering the C-arm source to iso-center distance (approximately 600mm), a 45° orbit of the C-arm results in approximately 430mm displacement of the X-ray source. Therefore, the aforementioned experiments all involve rotation as well as translation of the X-ray source.

The ranges selected for these experiments were solely defined by the limited capture volume or line of sight constraints of the tracking system, or the range of C-arm motion (e.g., swivel is only possible up to 10° for this C-arm).

2D/3D registration with automatic re-initialization

To evaluate the effect of re-initialization on intensity-based 2D/3D image registration, we attempt to register CT and C-arm images of a dry femur-pelvis phantom with the automatic re-initialization. In addition to re-initialization, the optimizer search space is limited according to the aforementioned tracking errors, i.e., the optimizer bounds are 10 and 20mm for <15° and >15° rotation of the C-arm, respectively. For speedup, the registration software is fully parallelized using GPU acceleration.

A CT scan is acquired from the femur-pelvis phantom that is encased in gelatin and partially covered with drapes (to simulate realistic surgical scenario). Next, the phantom is placed on

the surgical bed using laser guidance where the femur head is near the iso-center of the C-arm orbit. A C-arm X-ray image is then acquired and registered to the CT data using manual initialization. Thereafter, we rotate the C-arm within a range of 30° and collect multiple C-arm images while tracking the C-arm using vision-based methods in "Estimation of C-arm trajectory using visual tracking" section. For each C-arm image, the tracking outcome is used for re-initialization of the intensity-based 2D/3D registration. In Table 1, the registration outcome is presented for all successful registration attempts. We consider a registration attempt successful only if the mTRE error after registration is <2.5mm. mTRE is computed as the mean of the absolute Euclidean distances between radiolucent landmarks on the phantom in the CT coordinate frame. Registration using random initialization yielded a 23% success rate, where vision-based re-initialization yielded a 75% success rate. An example demonstrating the registration steps is presented in Fig. 5.

Discussion and conclusion

The marker-free pose-aware C-arm system proposed in this work is an RGBD enhanced C-arm system that uses vision-based techniques to track the relative displacement of the C-arm with respect to the surgical site. Before the vision-based tracking takes place, C-arm is registered to pre-interventional CT data using a single 2D/3D registration with manual interaction. Thereafter, the relative C-arm poses are estimated with respect to the initial C-arm pose and consequently to pre-interventional 3D data. An error in the initial registration of the C-arm to CT data will directly affect the accuracy of the vision-based tracking system. The fusion of 3D pre- and 2D intra-interventional data allows the surgeon to understand the relationship between the current state of surgery, the complex 3D structures, and the preoperative planning. This increases the surgeon's confidence, reduces the mental task load, and lowers the probability of a revision surgery.

Classically tracking the C-arm/patient using external optical trackers may provide better accuracy, but involves longer and more complex preparation, line of sight issues, and the invasive implantation of pins into bone, hampering surgical navigation from being widely adopted in orthopedics. In contrast to these systems, the pose-aware C-arm is noninvasive and requires no interaction time for manual re-initialization, and the workflow remains intact (manual initialization for authors as expert users takes between 30 and 60s). The registration step after a re-initialization takes <10 s.

The tracking accuracy is evaluated at various C-arm poses. The results in Fig. 4 indicate that an RGB+depth tracking method slightly outperforms a depth-only tracking system. The translation error for a wide range of displacement is 10–18mm, and the rotation error is below 2°. The reported mTRE values are only for the re-initialization step. The range of acceptable mTRE for clinical application depends on the registration method and its capture range. Surgical navigation can only be performed after the 2D/3D registration takes place.

Registration using the re-initialization has 75% success rate, where random initialization only yielded 23% success rate. Our method increases the likelihood of a successful registration, which therefore improves the practical applicability and usefulness of surgical navigation. A similar fiducial-less initialization method in [2] yielded only 68.8% success

rate, where in 95% of the cases the initialization error was below 19.5mm. Spine imaging is the focus of the 2D/3D registration proposed in [32], where DRRs are generated from each vertebra and used to generate Hough space parametrization of the imaging data. The success rate of the registration using Generalized Hough Transform is 95.73%. The interactive initialization method in [14] with average interaction time 132.1 ± 66.4s has an error of 7.4 ± 5.0 mm.

RGBD camera is mounted on the image intensifier that allows the X-ray source to remain under the surgical bed; hence, the workflow is not disrupted. Despite the benefits of mounting the camera near the detector, large distance between RGBD camera and the X-ray source (approximately 1000mm) may result in significant propagation of error. In other words, minute tracking error in the RGBD coordinate frame may produce large errors in the X-ray source coordinate frame.

The depth-only algorithm performs a global tracking and utilizes points from foreground as well as the background. Therefore, patient movement in a static background is not tracked. Moreover, this method requires a complex scene with dominant structures. Lastly, the tracking is only maintained when the C-arm is re-arranged slowly. On the other hand, RGBD SLAM uses color features and assigns a dense feature area in the image (e.g., surgical site) as the fore-ground, and discards the background. Therefore, RGBD SLAM accounts for patient rigid movement.

The errors from the vision-based tracking show that the translational components are subject to larger errors than the rotational components. The errors depend on the choice of the color features and the accuracy of depth data. The error in tracking the X-ray source using the RGBD SLAM is the result of unreliable color features, drift from the frame-to-frame tracking, large translations of the RGBD camera in the world coordinate frame, and the propagation of error due to the large distance between the X-ray and the RGBD imaging devices. Fast C-arm movement as well as the presence of reflective and dark objects in the surgical scene that reflect and absorb IR are factors that result in poor tracking. Tracking accuracy can be further improved by incorporating application specific details such as drape color and using model-based tracking methods based on the known representations of the surgical tools. Additionally, the drift caused by frame-to-frame tracking can be reduced using bundle adjustment. Other improvements can take place by using stereo RGBD cameras on the C-arm, or using a combined tracking algorithm with both RGBD SLAM and depth-only. Alternative feature detectors that provide better representations can be used to extract relevant information from the surgical scene. Redundancy tests may also be necessary to reject color features where the depth information is sparse or unreliable. Lastly, high-end RGBD imaging devices with more reliable depth information can significantly improve the tracking quality.

The tracking accuracy of the RGBD SLAM method presented in this paper is within the capture range of an intensity-based image registration and provides reliable registration. We believe that an RGBD enhanced C-arm system can contribute to automated fusion of pre- and intra-interventional data.

Acknowledgments

References

1. Bay, H., Tuytelaars, T., Van Gool, L. European conference on computer vision. Springer; Berlin: 2006. Surf: speeded up robust features; p. 404-417.

2. Van der Bom M, Bartels L, Gounis M, Homan R, Timmer J, Viergever M, Pluim J. Robust initialization of 2D–3D image registration using the projection-slice theorem and phase correlation. Med Phys. 2010; 37(4):1884–1892. [PubMed: 20443510]

3. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W. 2012 IEEE international conference on robotics and automation (ICRA). IEEE; 2012. An evaluation of the RGB-D SLAM system; p. 1691-1696.2012

4. Euler E, Heining S, Riquarts C, Mutschler W. C-arm-based three-dimensional navigation: a preliminary feasibility study. Comput Aided Surg. 2003; 8(1):35–41. [PubMed: 14708757]

5. Fischer M, Fuerst B, Lee SC, Fotouhi J, Habert S, Weidert S, Euler E, Osgood G, Navab N. Preclinical usability study of multiple augmented reality concepts for K-wire placement. Int J Comput Assist Radiol Surg. 2016; 11(6):1007–1014. [PubMed: 26995603]

6. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 1981; 24(6):381–395.

7. Fotouhi J, Fuerst B, Lee SC, Keicher M, Fischer M, Weidert S, Euler E, Navab N, Osgood G. Interventional 3D augmented reality for orthopedic and trauma surgery. 16th annual meeting of the international society for computer assisted orthopedic surgery (CAOS). 2016

8. Fotouhi J, Fuerst B, Wein W, Navab N. Can real-time rgbd enhance intraoperative cone-beam CT? Int J Comput Assist Radiol Surg. 2017; :1–9. DOI: 10.1007/s11548-017-1572-y

9. Fuerst, B., Fotouhi, J., Navab, N. International conference on medical image computing and computer-assisted intervention. Springer; Berlin: 2015. Vision-based intraoperative cone-beam ct stitching for non-overlapping volumes; p. 387-395.

10. Geiger, A., Moosmann, F., Car, Ö., Schuster, B. 2012 IEEE international conference on robotics and automation (ICRA). IEEE; 2012. Automatic camera and range sensor calibration using a single shot; p. 3936-3943.

11. Gendrin C, Furtado H, Weber C, Bloch C, Figl M, Pawiro SA, Bergmann H, Stock M, Fichtinger G, Georg D. Monitoring tumor motion by real time 2D/3D registration during radiotherapy. Radiother Oncol. 2012; 102(2):274–280. [PubMed: 21885144]

12. Glassner, AS. An introduction to ray tracing. Elsevier; Amsterdam: 1989.

13. Goitein M, Abrams M, Rowell D, Pollari H, Wiles J. Multi-dimensional treatment planning: Ii. beam's eye-view, back projection, and projection through ct sections. Int J Radiat Oncol Biol Phys. 1983; 9(6):789–797. [PubMed: 6863054]

14. Gong RH, Güler Ö, Kürklüoglu M, Lovejoy J, Yaniv Z. Interactive initialization of 2D/3D rigid registration. Med Phys. 2013; 40(12):121911. [PubMed: 24320522]

15. Grisetti, G., Kümmerle, R., Stachniss, C., Frese, U., Hertzberg, C. 2010 IEEE international conference on robotics and automation (ICRA). IEEE; 2010. Hierarchical optimization on manifolds for online 2D and 3D mapping; p. 273-278.

16. Jain, A., Fichtinger, G. International conference on medical image computing and computer-assisted intervention. Springer; Berlin: 2006. C-arm tracking and reconstruction without an external tracker; p. 494-502.2006

17. Knaan, D., Joskowicz, L. International conference on medical image computing and computer-assisted intervention. Springer; Berlin: 2003. Effective intensity-based 2D/3D rigid registration between fluoroscopic X-ray and CT; p. 351-358.

18. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W. 2011 IEEE international conference on robotics and automation (ICRA). IEEE; 2011. g2o: a general framework for graph optimization; p. 3607-3613.

19. Lee SC, Fuerst B, Fotouhi J, Fischer M, Osgood G, Navab N. Calibration of RGBD camera and cone-beam CT for 3D intra-operative mixed reality visualization. Int J Comput Assist Radiol Surg. 2016; 11(6):967–975. [PubMed: 27059022]

20. Lemieux L, Jagoe R, Fish D, Kitchen N, Thomas D. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. Med Phys. 1994; 21(11):1749–1760. [PubMed: 7891637]

21. Mitrovi U, Pernuš F, Likar B, Špiclin Ž. Simultaneous 3D–2D image registration and c-arm calibration: application to endovascular image-guided interventions. Med Phys. 2015; 42(11):6433–6447. [PubMed: 26520733]

22. Navab, N., Bani-Hashemi, A., Nadar, MS., Wiesent, K., Durlak, P., Brunner, T., Barth, K., Graumann, R. International conference on medical image computing and computer-assisted intervention. Springer; Berlin: 1998. 3D reconstruction from projection matrices in a c-arm based 3D-angiography system; p. 119-129.

23. Newcombe, RA., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, AJ., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A. 2011 10th IEEE international symposium on mixed and augmented reality (ISMAR). IEEE; 2011. Kinectfusion: real-time dense surface mapping and tracking; p. 127-136.

24. Otake Y, Armand M, Armiger RS, Kutzer MD, Basafa E, Kazanzides P, Taylor RH. 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. 2012; 31(4):948–962. [PubMed: 22113773]

25. Otake Y, Armand M, Sadowsky O, Armiger RS, Kutzer MD, Mears SC, Kazanzides P, Taylor RH. An image-guided femoroplasty system: development and initial cadaver studies. SPIE medical imaging. International Society for Optics and Photonics. 2010:76250P–76250P. 2010.

26. Otake Y, Schafer S, Stayman J, Zbijewski W, Kleinszig G, Graumann R, Khanna A, Siewerdsen J. Automatic localization of vertebral levels in X-ray fluoroscopy using 3D–2D registration: a tool to reduce wrong-site surgery. Phys Med Biol. 2012; 57(17):5485. [PubMed: 22864366]

27. Powell, MJ. Cambridge NA Report NA2009/06. Vol. 2009. University of Cambridge; Cambridge: 2009. The BOBYQA algorithm for bound constrained optimization without derivatives.

28. Rusinkiewicz, S., Levoy, M. Proceedings. Third international conference on 3-D digital imaging and modeling. Vol. 2001. IEEE; 2001. Efficient variants of the ICP algorithm; p. 145-152.

29. Sadowsky, O., Chintalapani, G., Taylor, RH. International conference on medical image computing and computer-assisted intervention. Springer; Berlin: 2007. Deformable 2D-3D registration of the pelvis with a limited field of view, using shape statistics; p. 519-526.2007

30. Uneri A, Otake Y, Wang A, Kleinszig G, Vogt S, Khanna AJ, Siewerdsen J. 3D–2D registration for surgical guidance: effect of projection view angles on registration accuracy. Phys Med Biol. 2013; 59(2):271. [PubMed: 24351769]

31. Varnavas A, Carrell T, Penney G. Increasing the automation of a 2D–3D registration system. IEEE Trans Med Imaging. 2013; 32(2):387–399. [PubMed: 23362246]

32. Varnavas A, Carrell T, Penney G. Fully automated 2D–3D registration and verification. Med Image Anal. 2015; 26(1):108–119. [PubMed: 26387052]

33. Wein, W. Master's thesis. Technische Universität; München: 2003. Intensity based rigid 2D-3D registration algorithms for radiation therapy.

34. Yang L, Wang J, Kobayashi E, Ando T, Yamashita H, Sakuma I, Chiba T. Image mapping of untracked free-hand endoscopic views to an ultrasound image-constructed 3D placenta model. Int J Med Robot Comput Assist Surg. 2015; 11(2):223–234.

35. Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell. 2000; 22(11):1330–1334.
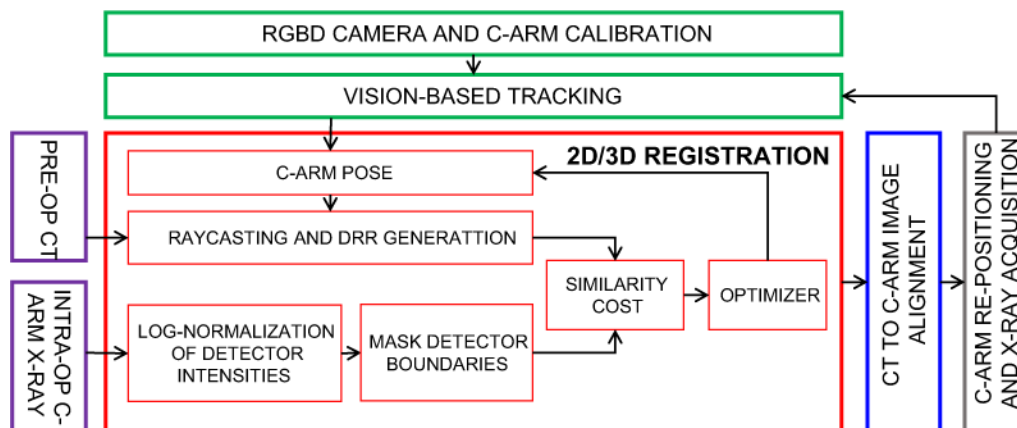
Fig. 1.
Workflow of a pose-aware C-arm system for intensity-based 2D/3D registration. Every C-arm image is globally aligned (initialization for 2D/3D registration) with pre-interventional CT based on vision-based tracking. The main contribution of this work is shown in *green*
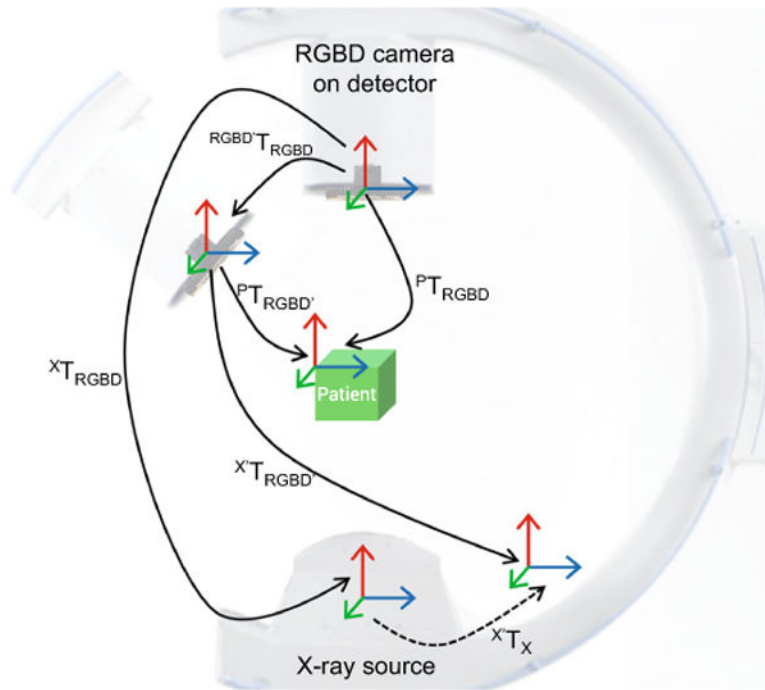
Fig. 2.

Relative displacement between camera poses $^{RGBD'}\mathbf{T}_{RGBD}$ is used for estimation of the relative displacement of the X-ray source $^{X'}\mathbf{T}_X$
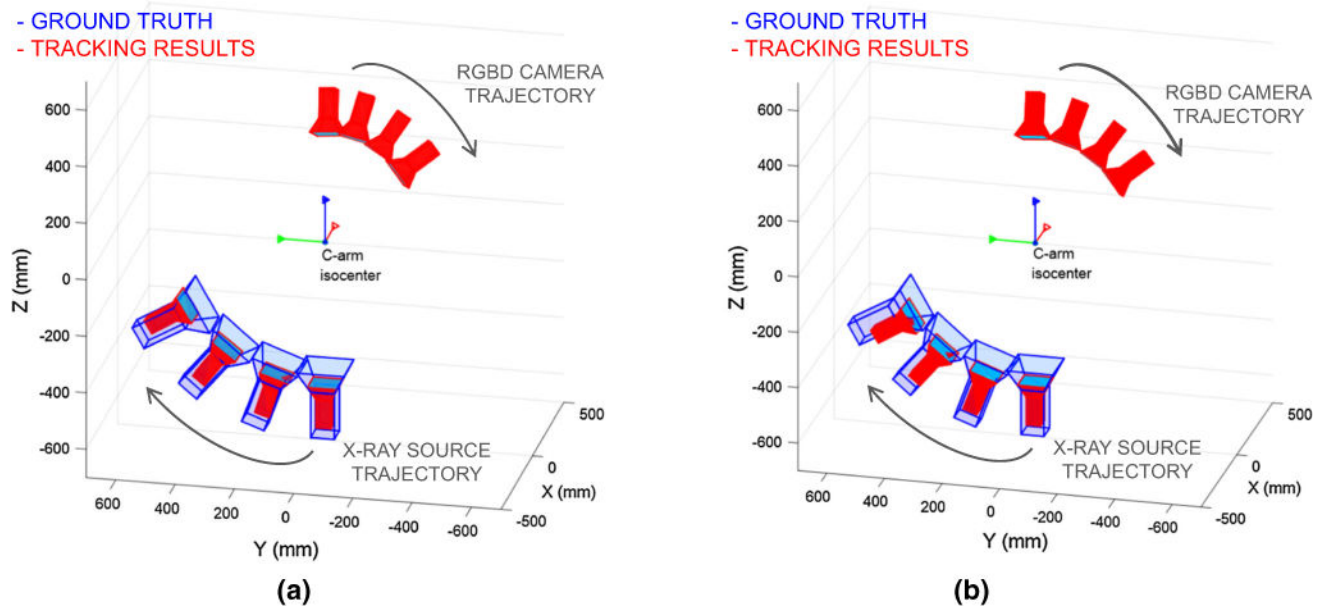
Fig. 3.
Pose of the X-ray source is tracked at different C-arm positions. Tracking based on **a** depth-only, and **b** an RGBD SLAM system is shown in *red*, and the tracking outcome based on the external tracking system is shown in *blue*. While drift (mistranslation) is observed in the tracking of the X-ray source, misrotation compared to the ground truth is minimal
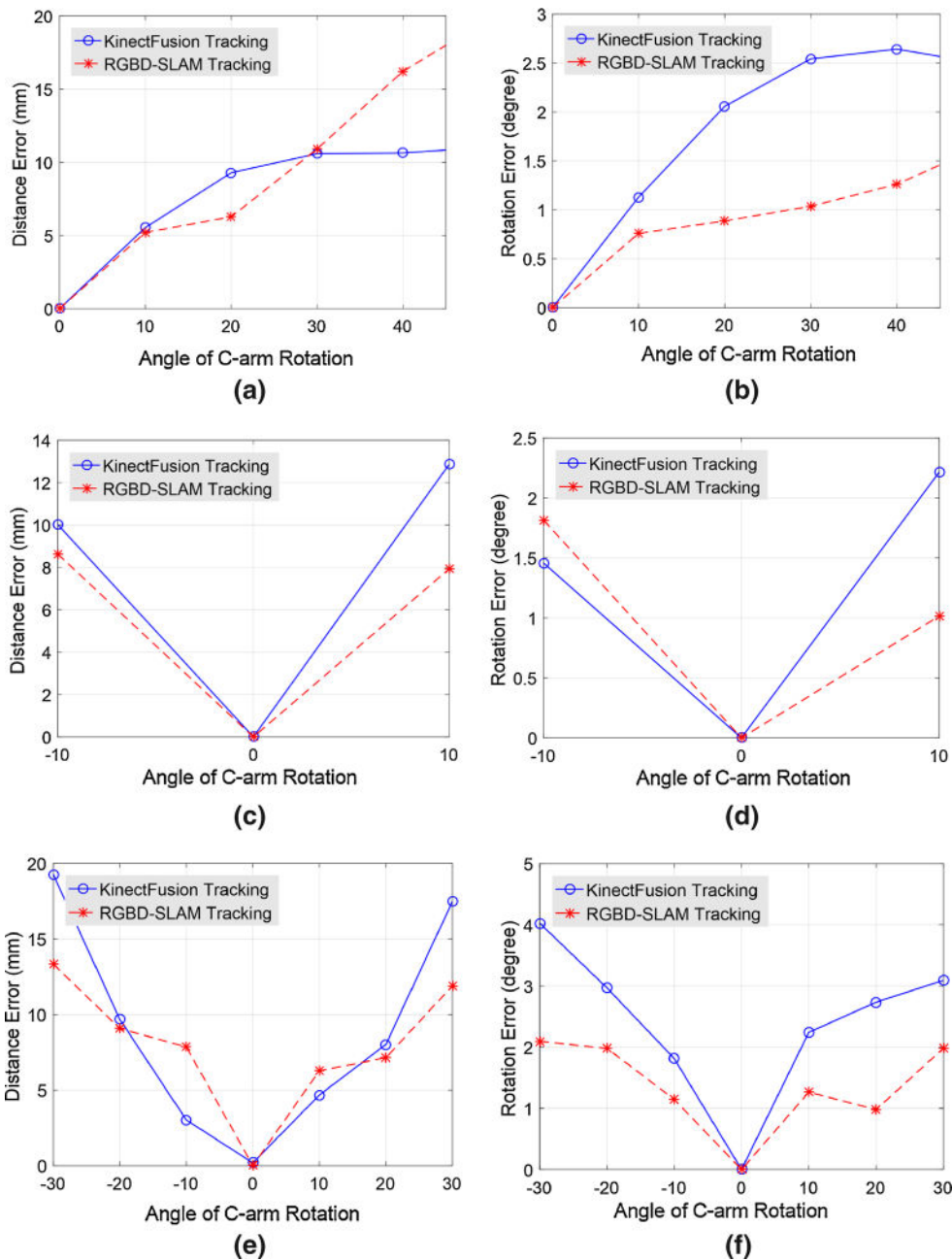
Fig. 4.
Tracking errors of X-ray source pose estimation with respect to ground truth are presented in Euclidean distance as well as rotational misalignment. The *plots* indicate the results when using depth-only data, and RGBD SLAM with RGB+depth information. **a**, **b** are the errors for 45° C-arm rotation in cranial/caudal direction, **c**, **d** correspond to 10° C-arm swivel, and **e**, **f** are errors for ±30° of C-arm rotation in oblique ± direction. Regarding the composition of errors, results show a relatively small rotational and large translational component. Note that here the C-arm extrinsics recovered from vision-based methods are compared to tracking results using an external optical tracking system. Therefore, any changes in the
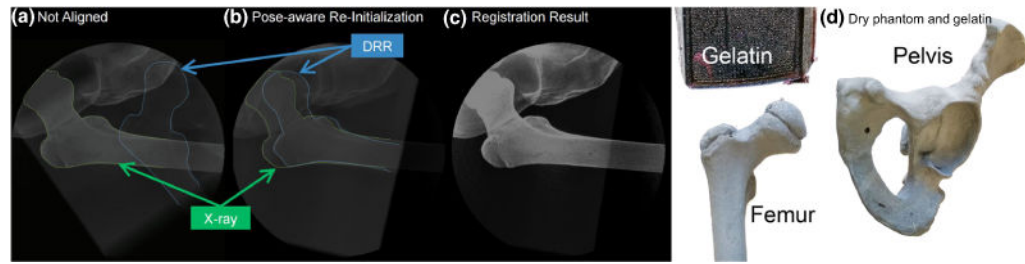
intrinsic parameters (due to mechanical deformation of the C-arm) do not contribute to the errors

Fig. 5.

DRR is overlaid with C-arm X-ray image, **a** before re-initialization, **b** after pose-aware re-initialization, and **c** after the 2D/3D registration. The pelvic-femur phantom and the gelatin are shown in **d**. During the experiment, the bone phantom is encased in gelatin and covered with drape to simulate a realistic surgical condition

Table 1

NCC similarity measures for the intensity-based 2D/3D registration are presented at two stages: (1) after re-initialization, (2) after registration

Tracking method	After re-initialization			After registration		
	NCC	SD (NCC)	mTRE	NCC	SD (NCC)	Success rate (%)
Depth-only	0.243	0.18	13.23	0.721	0.04	69
RGBD SLAM	0.310	0.24	11.81	0.749	0.06	75

The experiments are conducted using 13 arbitrary C-arm poses, and the mean and standard deviation (SD) of the similarity measure are presented in the table. Moreover, mTRE values are presented in mm after the re-initialization step. We consider a registration with mTRE <2.5mm as a successful attempt. The rate of success using each method is presented in the last column