CIS II: Paper Presentation Report

Jesse Hamilton

April 24, 2012

Group 3: X-ray Image Based Navigation for Hip Osteotomy

Team Member: Michael Van Maele

Mentors: Dr. Mehran Armand, Dr. Yoshito Otake, Ryan Murphy

Paper Reference:

*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. Medical Imaging, IEEE Transactions on 2012;31(4):948-962.*

Project Overview

Periacetabular osteotomy (PAO) is a joint reconstruction surgery for increasing femoral head coverage and improving stability in patients with developmental dysplasia of the hip (DDH). Our research mentors have previously developed a software package for geometrical and biomechanical planning of PAO called the Biomechanical Guidance System (BGS). One of its key features is intraoperative fragment tracking using a Polaris optical system.

The first aim of our project is to develop a protocol and software pipeline for an x-ray image-guided navigation system for performing PAO. C-arm imagers are more prevalent in hospitals than optical trackers, and surgeons already have familiarity with fluoroscopy procedures. Our second goal is to compare the proposed pipeline with the current optically navigated procedure.

Paper Selection

The paper by Dr. Otake (one of our mentors) et al. is extremely relevant to our project. The purpose of the article is to describe a framework for intraoperative 2D/3D image-based registration utilizing a conventional C-arm imager and graphics processing unit (GPU) acceleration. Broadly speaking, the proposed framework involves three steps: (1) collecting intraoperative x-ray images, (2) estimating relative pose between x-ray images using a radiopaque fiducial, and (3) estimating patient pose using 2D/3D registration.

We will be utilizing the same custom fiducial – called a fluoroscopic tracker (FTRAC) – as described in the article for estimating the relative pose between C-arm images. However, our pipeline utilizes a correspondenceless expectation-maximization algorithm to compute these relative poses, while the authors used a correspondence-based algorithm (POSIT – Pose from Orthography and Scaling with ITerations). In addition, the authors investigate how registration accuracy and computation time are influenced by different workflow components, such as different DRR generation algorithms, similarity measures, and optimization strategies. Our final pipeline will incorporate 2D/3D registration code provided by our mentors, and the current settings were chosen based on the results in this paper.

Paper Summary

*Background*

Image-based registration has several advantages over point-based registration. Surgeons do not need to physically contact landmark points on patient anatomy, and there is no need for manual segmentation of anatomical features. However, there are technical obstacles hindering routine use of 2D/3D registration in image-guided surgery (IGS). Unlike room-mounted imagers, mobile C-arms have reduced mechanical stability, which introduces uncertainty when estimating extrinsic and intrinsic parameters. In addition, creating digitally reconstructed radiographs (DRRs) is computationally expensive. The authors address the first problem by using an FTRAC and address the second by exploiting the parallel computing power of a GPU.

*Methods*

A diagram of the registration workflow is shown in Fig. 1. The next section will highlight the following key steps in the workflow: pre-operative calibration, C-arm pose estimation, image preprocessing, initialization, DRR generation, similarity measurement, and optimization.



Fig. 1. Diagram of registration workflow

Preoperative Calibration

In their experiments, the authors used both a conventional C-arm with an x-ray image intensifier (XRII) and a C-arm cone beam CT (CBCT) with a flat-panel detector. The former required distortion correction, while the latter did not. A conventional C-arm was used because it is more widespread in hospitals, while CBCT was used to validate their workflow by obtaining data without calibration errors.

C-arm Pose Estimation

The authors used a hybrid fiducial consisting of an optical marker attached to an FTRAC (Fig. 2). To determine the relative pose between x-ray images, the software first segmented four non-coplanar beads. Object pose with respect to the virtual x-ray source was computed using POSIT, and a projection of the CAD model of the FTRAC was created. The CAD projection and x-ray image were compared using mutual information, and this process was repeated using a Downhill Simplex Algorithm until it converged.



Fig. 2. Hybrid fiducial

Image Preprocessing

Both the preoperative CT and intraoperative x-ray images underwent preprocessing. The CT data was converted from Hounsfield Units to linear attenuation coefficients, and line integrals of the linear attenuation coefficient were computed from the x-ray projections.

DRR Generation

The authors investigated two algorithms for generating DRRs (Fig. 3) using a GPU. In ray-tracing with trilinear interpolation, intensity values are computed at fixed intervals along a ray from the virtual x-ray source to each virtual detector and then summed along the ray. Siddon’s ray-tracing method weights voxel intensity by the intersection length of the ray with each voxel. Siddon’s method is more accurate but also more computationally expensive.



Fig. 3 Comparison of the two DRR generation algorithms

Objective Function

The authors compared three similarity measures: mutual information (MI), normalized mutual information (NMI), and gradient information (GI). They described GI in detail because of its superior performance. Unlike MI and NMI, it does not penalize external objects that appear in one image but not the other, such as the FTRAC. Along with DRR generation, computation of the similarity measures was done on the GPU.

Optimization

Two optimization algorithms were tested: a downhill Simplex method and a stochastic search algorithm (CMA-ES, or Covariance Matrix Adaptation Evolution Strategy). Coarse-to-fine multiresolution with three downsampling steps (factors 4, 2, 1) was also tested.

Experimental Setup

Four experiments were designed in order of increasing complexity. The authors aimed to validate their proposed method with simpler data sets before testing it on images with noise, calibration error, and pose estimation error.

A similar evaluation protocol was used across all experiments. The authors tested 100 registration trials with randomly selected initial guesses resulting in initial mean target registration error (mTRE) from 0 to 10 mm, in 1 mm intervals (10 trials were done at each initial mTRE). The proposed workflow was evaluated on four criteria:

1. **mean target registration error** (mTRE), evaluated using virtual target points within a $50 mm^{3}$ grid centered on the femoral head.
2. **success rate**, defined (in most cases) as $mTRE<2.5 mm$
3. **number of function evaluations**
4. **computation time**

Experiment 1: Gold Standard

The first experiment used a gold standard data set containing x-ray and CT images of a cadaveric pig head. The purpose of this experiment was to compare the basic registration framework with results in published literature. Paired point registration was used as ground truth.

Experiment 2: Plastic Bone Phantom

In the second experiment, x-ray projections of a plastic bone phantom were acquired using CBCT with a flat-panel detector. Its purpose was to evaluate the proposed workflow with an ideal data set in the absence of noise from soft tissue and calibration errors. One to four images taken at evenly spaced angles were selected to test the registration accuracy. 3D/3D registration using mutual information was used as the ground truth.

Experiment 3: Cadaver with CBCT

The protocol for the third experiment was similar to the second experiment, except a cadaver specimen was used instead of a phantom. Its purpose was to evaluate the registration in the presence of noise from soft tissue.

Experiment 4: Cadaver with Conventional C-arm XRII

This section culminates with Experiment 4, which was designed to test the registration workflow in a realistic clinical setting with errors from XRII image distortion, uncertain geometric calibration, and pose-dependent changes of intrinsic C-arm parameters. X-ray images of a cadaveric femur were acquired using a conventional C-arm. Ground truth data was obtained by attaching the hybrid fiducial to the specimen, acquiring a preoperative CT, and performing rigid point-based registration between the x-ray images and CT volume. The C-arm was only calibrated once preoperatively, as would be the case in a real surgery.

Results/Discussion

In Experiment 1, the proposed method using gradient information achieved lower mean target registration and faster computation speed compared to published values. In Experiments 2-4, computation time was observed to increase linearly with the number of images used in registration (Fig. 4). In Experiment 2, 31% of trials failed when using a single image, while in both cadaver experiments all trials failed when using a single image. Importantly, the error in mTRE in Experiment 4 was twice as large as that in Experiment 3 $(2.04\pm 0.85mm $versus $0.99\pm 0.41 mm)$ . One source of increased error is likely from uncertainties in calibration and pose estimation from using a conventional C-arm. Because of the small standard deviation and large mean TRE in Experiment 4, the authors speculate that the ground truth registration may have been inaccurate. The optical/FTRAC fiducial may have shifted while the cadaver was in transport from the preoperative CT to the operating room.



Fig. 4. Computation time increased linearly with number of images.

The results of testing both DRR generation algorithms revealed that Siddon’s ray-tracing method was more accurate than ray-tracing with trilinear interpolation but required longer computation time. Decreasing the step length for trilinear interpolation improved the accuracy, with diminishing improvement below about 1 mm. Computation time decreased as the step size was made larger, with diminishing reductions above a step size of 1.5 mm. Subtracting a DRR processed with Siddon’s method from one created by trilinear interpolation revealed slight differences in image quality at a step length of 1.5 mm and striking differences at 3.0 mm (Fig. 5).



Fig. 5. Comparison of image quality between different DRR generating algorithms. The top row shows DRRs, and the bottom row shows difference images between trilinear interpolation and Siddon’s method.

Gradient information outperformed the two mutual information similarity measures. As pose parameters deviated from ground truth values, all three similarity measures showed a smooth profile with clear local maxima for Experiment 2 (plastic phantom) (Fig. 6). With Experiments 3 and 4 (cadaver), however, they displayed more jagged profiles due to noise from soft tissue. Compared to MI and NMI, GI had a more distinct peak in its profile in Experiment 4.

Fig. 6. Comparison of similarity metrics as translational and rotational pose parameters deviate from ground truth.

GI, MI, and NMI measure the similarity between one DRR and one x-ray image. The authors discuss possible ways to merge similarity measures when dealing with multiple image pairs. In this paper, they summed the similarity measure from each image pair. Another alternative not investigated here includes treating all the DRRs and x-ray projections as two composite images before computing the objective function. It is also possible to compute the similarity metric for one pair and rotate which images are used after each iteration.

The coarse-to-fine multilevel strategy was effective at increasing registration accuracy in Experiments 2 and 3 (Fig. 7). However, accuracy worsened in Experiment 4. The authors speculate that because of image noise and errors from pose estimation and calibration, errors produced in downsampled stages get magnified.



Fig. 7. Effect of the multiresolution approach

In conclusion, the authors claim that intraoperative registration using two x-ray images is feasible for image-guided surgery. The registration can be performed relatively quickly (in a matter of seconds), has reasonable accuracy and robustness, and is automated.

Personal Assessment

Overall, this article was clearly written and provided convincing evidence that it is feasible for 2D/3D registration to be used for image-guided orthopedic surgery. I liked how the authors provided explicit rationale behind the purpose of each experiment. The four experiments proceeded from testing simple, ideal data sets to noisier, more clinically realistic settings. Experiment 1 validated their new method by comparison with established literature. Experiment 2 tested the registration on a plastic phantom, with no confounding effects from soft tissue, image distortion, or pose estimation errors. Experiment 3 introduced noise from overlying tissue, while Experiment 4 replaced CBCT with a conventional C-arm.

The Methods section of this paper was well-presented. The authors clearly stated the two-pronged problem definition – estimating relative pose between x-ray images, and estimating relative pose between each x-ray image and the patient. They presented a high-level diagram of the registration workflow to guide readers through the rest of the section. Also, the Results section had an abundance of clearly annotated figures.

The paper did have a few minor weaknesses. A few technical details in the Methods could have been explained in more depth. For example, the authors did not mention exactly how the metallic beads in the FTRAC were segmented, and the description of C-arm pose estimation took several readings for me to grasp. In addition, the explanation of how the registration transformation was initialized was vague. I also was not satisfied with their explanation of the large mTRE in Experiment 4 being due to inaccuracies in the ground truth. It may have been useful to repeat that experiment, making sure that the fiducial remained fixed between acquiring CT and x-ray images. The authors mentioned how the coarse-to-fine multiresolution strategy actually worsened the accuracy in Experiment 4. They attributed this to noise from soft tissue, calibration errors, and pose estimation errors propagating from the low-resolution stages. Perhaps if they had an accurate ground truth, the multiresolution approach would have improved the registration.