A Novel Method of X-ray to CT Rigid Registration Using Scene Coordinate Regression

Yuichi Yoshii[1], Pragyan Shrestha[2], Chun Xie[2], Itaru Kitahara[2]
[1]Tokyo Medical University Ibaraki Medical Center, Ami, Japan. [2]Tsukuba University, Center for Computational Science, Tsukuba, Japan
yyoshii@tokyo-med.ac.jp

Disclosures: Yuichi Yoshii, Pragyan Shrestha, Xie Chun, Itaru Kitahara. No benefits in any form have been received from any commercial party related directly or indirectly to the subject of this study.

INTRODUCTION: Fluoroscopy is an essential tool for the current orthopaedic procedures. Due to overlapping anatomical structures in the fluoroscopic images, it is often difficult to correctly identify the 3D position and direction from solely the image. Therefore, registration of an intraoperatively acquired X-ray image to the preoperatively acquired CT scan is crucial in performing such procedures. Currently, the standard procedure for acquiring accurate registration involves attaching fiducial markers onto the patient and acquiring a preoperative CT scan. However, there exist two main issues with landmark estimation based methods; 1) need a sufficiently large number of landmarks in the CT image for the annotation, 2) possible failure of pose estimation in extreme views where projected landmarks are not visible or the number of visible landmarks is small. In this study, these issues are addressed by a novel method which introducing scene coordinates to establish dense 2D-3D correspondences. Specifically, the proposed method regresses the scene coordinates of the CT-scan model from corresponding X-ray images. A rigid transformation that aligns the CT-scan model to the image is then calculated by solving the perspective-n-point (PnP) problem with the random sample consensus (RANSAC) algorithm.

METHODS: The overview of the proposed registration pipeline is shown in Figure 1. The proposed method has four parts. First, the scene coordinate regression where a single view X-ray image is input to a U-Net model to obtain scene coordinates. Second, the PnP + RANSAC algorithm is used to solve for the pose of the captured X-ray system. Third, the segmentation of CT-scan volume to obtain a 3D model of bone. And fourth, the required rigid transformation of the 3D model from world coordinates to camera coordinates.

2D-3D Registration and Analysis

Scene coordinates are defined as the first point of intersection between a camera's back-projected rays and a 3D model in a world coordinate system. The same concept is adapted for X-ray images and its underlying 3D model obtained from CT-scan. Estimating the scene coordinates given an X-ray image results in dense implicit 2D-3D correspondences between the image points and the scene coordinates. After filtering out the non-intersecting scene coordinates using estimated standard deviation, the PnP problem is solved using RANSAC. An example of a successful registration is shown in Figure 2. To properly evaluate the proposed method, a dataset that contains six annotated CT scans each with several registered real X-ray images was used. The annotations include 14 landmarks and seven segmentation labels. Left anterior oblique/right anterior oblique of [-45, 45] degrees respectively were samples at one degree intervals. Random off-set was added in each direction. The CT scans of pelvic bones from six cadaveric specimens were evaluated. In total, 8100 simulated X-rays were generated, of which, 5184 images were randomly assigned as the training set, 1286 for the validation set, and the remaining 1620 for the test set. The U-Net was implemented in Pytorch 1.13.0. An image size of 512×512 is used for input as well as output scene coordinates. The output channel size is 8 (i.e., 3 for scene coordinates and 1 for standard deviation, multiplied by 2 for entry and exit points). The model was trained using Adam with a constant learning rate of 0.0001 and batch size of 16. Online data augmentation with a probability of 0.5 was applied for domain randomization. It included random invert, color jitter with brightness and contrast parameters each set to 1, and random erasing.

The proposed method was compared against two other baseline methods: PoseNet and DFLNet. PoseNet was implemented using ResNet-50 as the backbone for the feature extractor and trained using geometric loss. DFLNet uses the same architecture as the proposed method however the last layer regresses 14 heatmaps of the landmarks instead of scene coordinates. The mean target registration error and gross failure rate (GRF) was used as the evaluation metric to compare with the baselines. The differences of the GRF for each method were analyzed with one-way analysis of variance. P values of less than 0.05 were considered significant.

RESULTS: The mean target registration errors for the PoseNet, DFLNet, and proposed method were 7.6+/-1.6 mm, 49.6+/-69.3 mm, and 3.9+/-1.8mm, respectively. The mean GRFs for the PoseNet, DFLNet, and proposed method were 31.6+/-12.6%, 44.9+/-14.9%, and 12.2+/-7.8%, respectively. The GFR of proposed method was significantly smaller than those of other baseline methods (P<0.05). For most of the specimens, the proposed method could retain the GRF below 20% whereas PoseNet and DFLNet fail to register with more than 20% GRF in most cases. For PoseNet, this is because the network cannot reason about the spatial structure and its local relation to the image patches. For DFLNet, this is expected due to the visibility issue of landmark points that were mostly located in the pubic region of the pelvis.

DISCUSSION: This study presented a novel method of scene coordinate regression-based approach for the X-ray to CT-scan model registration. The proposed method does not require labeling of anatomical landmarks and is effective in extreme view angles. It was suggested that the proposed method could perform well even under partially visible structures and extreme view angles, compared to direct pose regression and landmark estimation based methods. Testing the model trained solely on simulated X-ray images, on real X-ray images did not result in catastrophic failure, instead, the results were positive for instantiating further refinement steps.

SIGNIFICANCE/CLINICAL RELEVANCE: A novel method of X-ray to CT rigid registration using scene coordinate regression was developed. This method will be useful for the orthopaedic procedures which using fluoroscopy.

ACKNOWLEDGEMENTS: This study is supported by Japan Society for the Promotion of Science (KAKENHI 23K08618) and the General Insurance Association of Japan.

FIGURES: