2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

3D Reconstruction of Tibia and Fibula using One General Model and Two X-ray Images

Kai Pan[1,*], Shuai Zhang[1,*], Liang Zhao[1], Shoudong Huang[1], Yanhao Zhang[2], Hua Wang[3] and Qi Luo[3]

***Abstract*— The 3D reconstruction of patient specific bone models plays a crucial role in orthopaedic surgery for clinical evaluation, surgical planning and precise implant design or selection. This paper considers the problem of reconstructing a patient-specific 3D tibia and fibula model from only two 2D X-ray images and one 3D general model segmented from the lower leg CT scans of one randomly selected patient. Currently, the bone 3D reconstruction mainly relies on computed tomography (CT) and magnetic resonance imaging (MRI) scanning-based mode segmentation which result in high radiation exposure or expensive costs. While, the proposed algorithm can accurately and efficiently deform a 3D general model to achieve a patient-specific 3D model that matches the patient's tibia and fibula projections in two 2D X-rays. The algorithm undergoes a preliminary deformation, 2D contour registration, and optimisation based on the deformation graph that represents the shape deformation of models. Evaluations using simulations, cadaver and in-vivo experiments demonstrate that the proposed algorithm can effectively reconstruct the patient's 3D tibia and fibula surface model with high accuracy.**

I. INTRODUCTION

Total knee arthroplasty (TKA) and unicompartmental knee arthroplasty (UKA) are consistently successful surgeries to effectively help patients alleviate pain and recover the function caused by knee related problems [1][2]. In the last few years, the number of operations of TKA and UKA have dramatically increased. For example, 374,833 TKA procedures in China were performed in 2019 [3], and it is predicted that the number of TKA procedures in the United States will increase to 1.26 million in 2030 [4].

To successfully perform the TKA and UKA surgeries, obtaining the accurate patient specific 3D tibia and fibula models is an important step for the clinical diagnosis, precise knee implants fitting, surgical planning and guidance, and postoperative evaluation. Currently, computed tomography (CT) and magnetic resonance imaging (MRI) are popular methods to obtain a 3D model because they both can generate a clear internal body structure. However, the radiation dose used for CT is much higher than X-ray images [5], which will inevitably affect patients' health. In addition, taking CT or MRI is a complex process requiring sophisticated devices and skillful manipulation, which will generate a high cost for both hospitals and patients. On the other hand, taking two X-ray images is a routine step before the operation. Therefore, 3D tibia and fibula model reconstruction technique by using X-ray images is very beneficial because patients can suffer less radiation dosage and financial pressure compared with CT or MRI scans.

Previous approaches mainly focused on the 3D reconstruction of bones using statistical shapes, intensity models and X-ray images. For example, [6] and [7] used two (or more) calibrated X-ray images and one statistical shape model (SSM) to reconstruct a patient specific 3D knee joint surface model. In detail, the patient specific knee joint model is estimated by optimising the SSM parameters that maximise the similarity between the bone contours extracted from the SSM projections and the bone contours extracted from the X-ray images. Later, [8] reconstructed the proximal femur bone (with interior density) using one statistical shape and intensity models (SSIM). The femur bone model was reconstructed by optimising the deformable SSIM parameters that maximise the intensity similarity or the intensity gradient similarity between the input X-ray images and the digitally reconstructed radiographs (DRRs) generated from the deformable model.

However, the construction of SSM and SSIM requires a large number of point-distributed knee joint model samples, which is not suitable for cases when the number of available models is limited. Furthermore, the numerical optimisation of the deformable SSM or SSIM parameters is computational expensive and requires very accurate initial X-ray frame poses [9].

Recently, deep learning approaches [10], [11] and [12] are proposed to learn bone shapes from multiple X-ray images. [10] reconstructed the lower leg bones using a SSM and the bone contour features extracted by a convolutional neural network. [11] and [12] used paired DRRs and CT scans to train a network that can predict 3D knee bones from two X-ray scans. However, a large number of DRRs, CT scans and also actual X-ray images are needed to train the networks.

Very recently, the 3D reconstruction technique using 2D images has been proposed and developed for aorta, not bones, which used a few 2D images to generate a 3D model [13] [14]. This technique is significant, especially when the 3D deformation information is difficult to obtain. However, the accuracy of this method still needs to be improved when the difference between the image observations and the deformation model is large.

In this paper, we proposed a method which only uses

* The first two authors have equal contributions.

[1]Kai Pan, Shuai Zhang, Liang Zhao and Shoudong Huang are with Robotics Institute, Faculty of Engineering and Information Technology, University of Technology Sydney (e-mails: Kai.Pan@student.uts.edu.au; Shuai.Zhang@student.uts.edu.au; Liang.Zhao@uts.edu.au; Shoudong.Huang@uts.edu.au).

[2]Yanhao Zhang is with School of Computing, Australian National University. Email: Yanhao.Zhang@anu.edu.au

[3]Hua Wang and Qi Luo are with Osteoarthropathy Surgery Department, Shenzhen People's Hospital, Shenzhen 518020, China (e-mail: wanghuayisheng@sina.com; l67q82@gmail.com).

one general 3D model and two X-ray images to reconstruct a patient-specific tibia (with fibula) model. Different from the reconstruction of SSM model/ SSIM models, needing a large dataset of knee joint model samples, the general model used in this work was segmented from a set of leg CT scans of a randomly selected patient. Besides, it does not need to iteratively generate DRRs and compare the intensity distribution or gradient, which significantly reduces the computational cost. Instead, this paper uses an algorithm similar to iterative closest point (ICP) [15] for 2D to 3D correspondence matching and embedded deformation (ED) algorithm [16] for estimating the deformation parameters. Experimental results using simulations, cadaver, and in-vivo datasets clearly demonstrated the efficiency and accuracy of the proposed approach and the advantages over some existing techniques [6][13].

II. PROBLEM STATEMENT

The problem considered in this paper is as follows. **Based on one 3D general model of tibia (with fibula), reconstructing a patient's 3D tibia and fibula with only two different views of X-ray images taken from the patient lower leg by a digital radiography machine.** Fig. 1 illustrates the input and output of the entire reconstruction system, where the two X-ray images are captured for the patient's lower leg from the coronal and the sagittal views, respectively.

To make our algorithm compatible with different X-ray devices, both perspective projection and orthogonal projection have been considered in this research. The perspective projection model can be described by

$$\boldsymbol{p}_{pers.} = UK\frac{1}{z}(R_{pers.}\boldsymbol{P} + \boldsymbol{t}_{pers.}) + \boldsymbol{\epsilon}_{pers.}, \qquad (1)$$

where $\boldsymbol{p}_{pers.} \in \mathbb{R}^2$ in pixel coordinate is the projection of a vertex $\boldsymbol{P} \in \mathbb{R}^3$ in the 3D model. $U = [I_2, \mathbf{0}_{2\times 1}]$ is used to take upper two rows of the calculated matrix, and K is the intrinsic matrix of the X-ray camera. $R_{pers.} \in SO(3)$ and $\boldsymbol{t}_{pers.} \in \mathbb{R}^3$ are the rotation matrix and translation vector that transform the vertices from the model frame to the X-ray frame, and $z \in \mathbb{R}$ is the z-axis value of vertex $\boldsymbol{P}$ in the camera frame. $\boldsymbol{\epsilon}_{pers.} \in \mathbb{R}^2$ is the zero-mean Gaussian noise with covariance matrix Ω.

Similarly, the orthogonal projection model can be expressed by

$$\boldsymbol{p}_{orth.} = sUR_{orth.}(\boldsymbol{P} + \boldsymbol{t}_{orth.}) + \boldsymbol{\epsilon}_{orth.}, \qquad (2)$$

where $\{R_{orth.}, \boldsymbol{t}_{orth.}\}$ denotes the relative pose of the digital radiography machine w.r.t. the 3D model, and $s > 0$ is the scale factor between the model and image coordinates.

In this paper, we assume $R_{pers.}, R_{orth.}, \boldsymbol{t}_{pers.}, \boldsymbol{t}_{orth.}$ are all known since the poses can be recovered from the header files of the X-ray images. Then, we can project the general 3D model using (1) or (2) to obtain two different views of simulated X-ray images (model projections) and extract the bone contours from the simulated X-ray images using the alpha-shape algorithm [17]. After that, the patient specific model can be estimated by deforming the general 3D model such that the bone contours from the deformed 3D model projections match the contours in the actual X-ray images. Section III will provide the details of the proposed method.

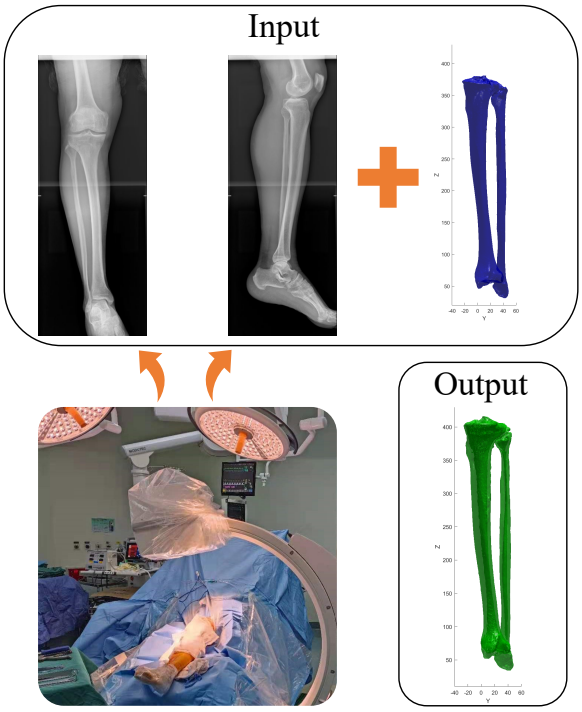


Fig. 1: The input and output of the reconstruction system. The left X-ray image is in the coronal plane, and the right X-ray image is in the sagittal plane, which are captured by a digital radiography machine; the blue model is a general 3D model segmented from a set of lower leg CT scans of a randomly selected patient, and the output (in green) is the reconstructed patient-specific 3D model.

III. METHODOLOGY

The proposed method sequentially follows four main steps, preliminary deformation, 2D-3D registration, deformation optimisation problem formulation, and solving the optimisation.

A. Preliminary Deformation

It is noted that each person has a different size (height and width) and shape of the tibia and fibula. If the size difference is large, it will cause difficulties in point matching during the 2D-3D registration (Section III-B). To avoid this issue, we first perform a preliminary deformation on the general 3D model based on the estimated difference in size.

The difference in size can be estimated by comparing the simulated X-ray images of the deformed model with the actual X-ray images. In detail, we can calculate two ratios r_h and r_w (one in height, one in width) as follows

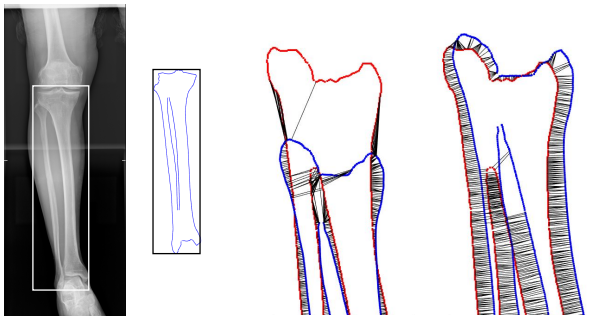
$$r_h = \frac{1}{2}\left(\frac{h^1_{box}}{\hat{h}^1_{box}} + \frac{h^2_{box}}{\hat{h}^2_{box}}\right), \quad r_w = \frac{1}{2}\left(\frac{w^1_{box}}{\hat{w}^1_{box}} + \frac{w^2_{box}}{\hat{w}^2_{box}}\right) \qquad (3)$$

where h^k_{box} and w^k_{box} are the height and width of the bounding box of the bone in the actual k-th X-ray ($k = 1, 2$), $\hat{h}^k_{box}$ and $\hat{w}^k_{box}$ are the height and width of the bounding box of the bone in the simulated k-th X-ray images, respectively. Fig. 2 (a) shows the example bounding boxes of the actual X-ray images and the simulated X-ray images.

Then, we can deform the general model using the estimated ratios. Specifically, each vertex $\boldsymbol{P}_i \in \mathbb{R}^3$ in the deformed model can be calculated by:

$$\boldsymbol{P}_i = \text{diag}(r_w, r_h, r_w)(\breve{\boldsymbol{P}}_i - \bar{\boldsymbol{P}}) + \bar{\boldsymbol{P}} \tag{4}$$

where $\breve{\boldsymbol{P}}_i \in \mathbb{R}^3$ is the i-th vertice in the general model, and $\bar{\boldsymbol{P}} \in \mathbb{R}^3$ is the geometric centre of the general model.



(a) (b)

Fig. 2: Preliminary deformation and its importance. (a) Example of the bounding box in the actual X-ray frame (left) and the simulated X-ray image (right); (b) The false contour correspondences before the preliminary deformation (left), the correct correspondences after the preliminary deformation (right).

Fig. 2 (b) shows that after the preliminary deformation, the bone contour points in the preliminary deformed model projection can be easily match to the bone contour features in the actual X-ray image. While without the preliminary deformation, it is very difficult to obtain the correct matches.

B. 2D-3D Registration

The 2D-3D registration aims to create the association between the contour points on the actual X-ray frames and the 3D vertices in the preliminary deformed model.

In this step, the projection of the preliminary deformed model will be first created by (1) or (2). Then we perform the points matching between the bone contour of the model projection and the bone contour of the corresponding actual X-ray image.

Suppose $\hat{\boldsymbol{p}} = \{\hat{\boldsymbol{p}}_1, ..., \hat{\boldsymbol{p}}_{N_{\hat{\boldsymbol{p}}}}\}$ represent the bone contour features of the actual X-ray images, $\tilde{\boldsymbol{p}} = \{\tilde{\boldsymbol{p}}_1, ..., \tilde{\boldsymbol{p}}_{N_{\tilde{\boldsymbol{p}}}}\}$ and $\tilde{\boldsymbol{P}} = \{\tilde{\boldsymbol{P}}_1, ..., \tilde{\boldsymbol{P}}_{N_{\tilde{\boldsymbol{p}}}}\}$ represent the contour points of the model projection in 2D frame and 3D model frame. Then, the point correspondence of $\hat{\boldsymbol{p}}_i \in \hat{\boldsymbol{p}}, i \in \{1, ..., N_{\hat{\boldsymbol{p}}}\}$ from $\tilde{\boldsymbol{p}}$ will be obtained by first evaluating the distances and the normal vectors to obtain the set

$$\begin{aligned}\tilde{\boldsymbol{p}}_i = \{\tilde{\boldsymbol{p}}_j | \tilde{\boldsymbol{p}}_j \in \tilde{\boldsymbol{p}}, \quad & d(\hat{\boldsymbol{p}}_i, \tilde{\boldsymbol{p}}_j) < d_{th}, \\ & \theta(\boldsymbol{n}(\hat{\boldsymbol{p}}_i), \boldsymbol{n}(\tilde{\boldsymbol{p}}_j)) < \theta_{th}\}\end{aligned} \tag{5}$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance between two points, $\boldsymbol{n}(\cdot)$ denotes a point's normal vector, $\theta(\cdot, \cdot)$ denotes the angle between two normal vectors. d_{th} and θ_{th} are the distance threshold and angle threshold. If $\tilde{\boldsymbol{p}}_i$ is not empty, the corresponding 3D model point $\tilde{\boldsymbol{P}}_i$ with the smallest $d(\hat{\boldsymbol{p}}_i, \tilde{\boldsymbol{p}}_j)$ is selected as the matched correspondence for $\hat{\boldsymbol{p}}_i$.

C. Optimisation Problem Formulation

This section describes the deformation estimation problem formulation using the preliminary deformed model and the 2D-3D points correspondences obtained from Section III-B. The main idea is to calculate the deformation graph using the ED algorithm [16], which aims to perform an overall deformation by using all the local deformations from the m ED nodes which are uniformly sampled from the preliminary deformed model vertices. Each vertex after embedded deformation $\breve{\boldsymbol{P}}_{\boldsymbol{i}} \in \mathbb{R}^3$ can be calculated by

$$\begin{aligned}\breve{\boldsymbol{P}}_{\boldsymbol{i}} &= \sum_{j=1}^{N} \omega_j(\boldsymbol{P}_i)[A_j(\boldsymbol{P}_i - \boldsymbol{g}_j) + \boldsymbol{g}_j + \boldsymbol{t}_j], \\ \omega_j(\boldsymbol{P}_i) &= (1 - ||\boldsymbol{P}_i - \boldsymbol{g}_j||/d_j)/n_j,\end{aligned} \tag{6}$$

where $\boldsymbol{P}_i$ is the vertex in the 3D pre-deformed model, $\boldsymbol{g}_j \in \mathbb{R}^3$ represents the 3D coordinates of the j-th ED node, $A_j \in \mathbb{R}^{3\times3}$ and $\boldsymbol{t}_j \in \mathbb{R}^3$ are the affine transformation matrix and translation vector describing the deformation parameters of ED node $\boldsymbol{g}_j$. $\omega_j(\boldsymbol{P}_i)$ is the affecting weight of nearest points, which depends on d_j (the distance between this point and the $(N+1)$-th nearest node), and n_j is functioned as the normalisation factor.

The affine transformation matrix A_j and translation vector $\boldsymbol{t}_j$ can be determined by minimising the energy function

$$E = \omega_{rot}E_{rot} + \omega_{reg}E_{reg} + \omega_{ob}E_{ob} \tag{7}$$

which combined the three terms E_{rot}, E_{reg}, E_{ob} with weights $\omega_{rot}, \omega_{reg}, \omega_{ob}$. The rotation term E_{rot} is used to make the affine matrices close to rotations, and the regularisation term E_{reg} is used to make the deformation smooth by reducing the divergence between one node itself and its neighbouring nodes. These two terms are the same as those in the original ED algorithm [16].

The aim of our new defined observation term E_{ob} is to make the model projection contours align with the contours extracted from the real X-ray images, so the observation term is the sum of squared distance between all 2D-3D corresponding point. For the perspective projection, the observation term is formulated as:

$$E_{ob} = \sum_{k=1}^{2} \sum_{i \in \mathbb{N}(k)} ||UK\frac{1}{z}(R_k\breve{\boldsymbol{P}}_{ki} + \boldsymbol{t}_k) - \hat{\boldsymbol{p}}_{ki}||^2, \tag{8}$$

where k is the index of images, $\hat{\boldsymbol{p}}_{ki}$ is the feature on the X-ray images, and $\breve{\boldsymbol{P}}_{ki}$ is vertex in deformed 3D model which is calculated by (6). And for the orthogonal projection, the observation term is formulated as:

$$E_{ob} = \sum_{k=1}^{2} \sum_{i \in \mathbb{N}(k)} ||sUR_k(\breve{\boldsymbol{P}}_{ki} + \boldsymbol{t}_k) - \hat{\boldsymbol{p}}_{ki}||^2. \tag{9}$$

D. Iterative Solver

Algorithm 1: Tibia and Fibula Reconstruction

Input: One general tibia (with fibula) model, contours from two X-ray images, poses of image frames.
Output: Personalised 3D tibial and fibular model.
Initialisation: Performing preliminary deformation, ED nodes sampling and initialisation.

while *Algorithm not converged* **do**
  Step 2: 2D-3D registration
  Extract the model projection contours;
  Calculate normal vectors of all features $\tilde{\boldsymbol{p}}$ and $\hat{\boldsymbol{p}}$;
  while *each X-ray image* **do**
    while *each feature in the* $\hat{\boldsymbol{p}}$ **do**
      Find point correspondences by (5);
    end
  end
  Step 3: Optimisation
  Initialise state vector $\boldsymbol{X}$;
  Calculate $\boldsymbol{f}(\boldsymbol{X})$, $\boldsymbol{J}(\boldsymbol{X})$ and $\Delta\boldsymbol{X}$;
  if $F(\boldsymbol{X}+\Delta\boldsymbol{X}) < F(\boldsymbol{X})$ **then**
    Update state vector $\boldsymbol{X} = \boldsymbol{X} + \Delta\boldsymbol{X}$;
  else
    if $F(\boldsymbol{X}+\Delta\boldsymbol{X}) \geq F(\boldsymbol{X})$ **then**
      Perform line search for damping factor α;
      Update state vector $\boldsymbol{X} = \boldsymbol{X} + \alpha\Delta\boldsymbol{X}$;
    end
  end
  Step 4: Update model vertices by (6) and do the updated model projections.
end

Minimising the energy function (7) is a non-linear least squares problem and a variation of Gauss-Newton method is used to obtain the solution. If we denote the energy function E as $F(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{f}(\boldsymbol{X})$, where $\boldsymbol{X}$ is the state vector including all the affine transformation matrix A_j and translation vector $\boldsymbol{t}_j$. In each iteration of the Gauss-Newton method, the step change $\Delta\boldsymbol{X}$ is computed by

$$(\boldsymbol{J}^\top \boldsymbol{J})\Delta\boldsymbol{X} + \boldsymbol{J}^\top \boldsymbol{f} = 0, \tag{10}$$

where $\boldsymbol{J}$ is the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{X})$. More details are shown in Algorithm 1.

IV. RESULTS & DISCUSSION

The robustness and accuracy of the proposed framework are assessed by ten simulation datasets and also validated by one cadaver experiment and four in-vivo experiments, which covered both perspective and orthogonal projections.

A. Error Measurement

First, the average point to plane distance is used to measure the error between the final deformed model and the ground truth. Second, for the clinical purpose, we have also compared the mechanical axes (an essential reference in pre-operative planning and marked by orthopaedic surgeons) between the deformed model and its ground truth from the coronal plane and the saggital plane, respectively.

B. Simulations

In all simulations, the same general 3D model, ground truth model (used for evaluation and segmented from another patient's lower leg CT scans) and two different views of grayscale images obtained from the ground truth model projections are used.

To simulate in-vivo scenarios and test the robustness of the proposed framework, in each simulation dataset, 5 different levels of zeros-mean Gaussian noise with standard deviation of $\{0.2, 0.5, 1, 1.5, 2\}$ pixels are added to the bone contour observations extracted from the grayscale images.

In each simulation dataset, the Euler angles of the frame poses for the first and second view of model projections are $\{\alpha_1 = 173^\circ,\ \beta_1 = 55^\circ,\ \gamma_1 = 5^\circ\}$ and $\{\alpha_2 = 173^\circ,\ \beta_2 = -25^\circ,\ \gamma_2 = 5^\circ\}$, respectively. In the 2D-3D registration step, d_{th} is set to 80 pixels and θ_{th} is set to $\frac{\pi}{3}$ rad. The number of nearest nodes N is set to 6 in the ED process, and the weight ratio between ω_{rot}, ω_{reg} and ω_{obs} in (7) is 1:10:100.

TABLE I: Mean reconstruction errors of the simulations (1^{st}-5^{th}: Perspective; 6^{th}-10^{th}: Orthogonal)

Noise (pixel)	0	0.2	0.5	1.0	1.5	2.0
1^{st} Set (mm)	1.131	1.155	1.120	1.505	1.417	1.900
2^{nd} Set (mm)	1.131	1.273	1.167	1.441	1.462	1.640
3^{rd} Set (mm)	1.131	1.122	1.236	1.381	1.543	1.542
4^{th} Set (mm)	1.131	1.125	1.293	1.266	1.599	1.520
5^{th} Set (mm)	1.131	1.548	1.236	1.380	1.387	1.421
Average (mm)	1.131	1.245	1.210	1.395	1.482	1.605
6^{th} Set (mm)	1.016	1.049	1.072	1.191	1.392	1.454
7^{th} Set (mm)	1.016	1.031	1.161	1.420	1.286	1.426
8^{th} Set (mm)	1.016	1.050	1.145	1.041	1.099	1.528
9^{th} Set (mm)	1.016	1.026	1.124	1.238	1.165	1.263
10^{th} Set (mm)	1.016	1.044	1.189	1.119	1.407	1.306
Average (mm)	1.016	1.040	1.138	1.202	1.270	1.395

Table I showed the mean reconstruction errors (unit in mm) between each estimated model and its corresponding ground truth model with the increasing noise levels. The error is positively correlated to the noise level after calculating their average error, which starts from 1.131 mm (zero noise) to 1.605 mm (2 pixels noise) in perspective projection, and from 1.016 mm (zero noise) to 1.395 mm (2 pixels noise) in orthogonal projection. However, the reconstructions still have a high accuracy even with 2 pixel noise since the proposed registration approach can provide robust correspondence and the regularization term in (7) contributes to the robustness for noisy observation.

Fig. 3 shows the deformation process at three different stages for one simulation dataset using the fifth level of noise. At the initial stage, the general model is much different from the ground truth, especially in length. After the preliminary deformation, the model has the similar size (length and width) comparing to the ground truth, but there is still a large difference in the model surface. For the final deformed

model, its model surface and ground truth surface overlapped and are consistent, which shows high reconstruction accuracy of the proposed algorithm.

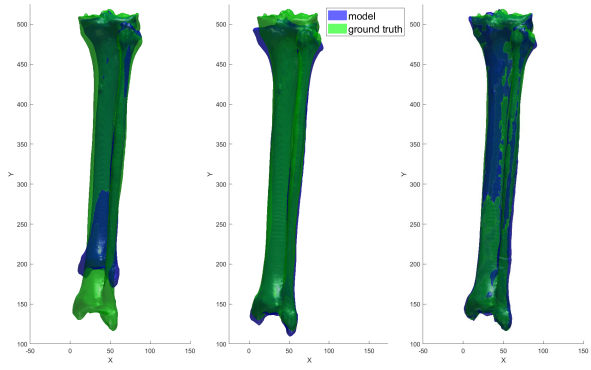


Fig. 3: Comparison of the ground truth (green) with the general model (left, blue), the preliminary deformed model (middle, blue), and the final model obtained from our algorithm (right, blue), for simulation data with the fifth level of noise added.

Fig. 4 compared the model projection contours and the observation contours at different stages. As we can see from both the coronal and sagittal views, the model projection contour at the initial stage has large differences in length compared the contour observation. Although the preliminary deformed model has the same length and width, its model projection contour has obvious misalignment with the contour observation. After the final deformation, the optimised model projection contour and its contour observation are overlapped in a good quality.

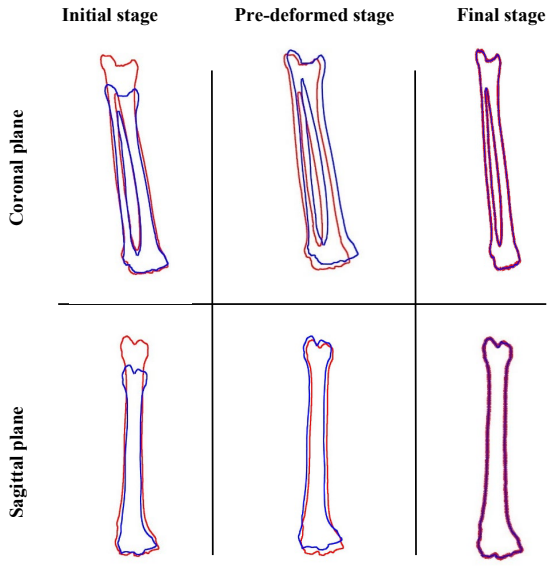


Fig. 4: Comparison between model projections (blue) and observations (red) at different stages in simulation.

Fig. 5 indicated the comparison between the estimated mechanical axis and the ground truth from the simulation with the fifth level of noise. The intersected angle by the mechanical axes between the reconstructed model and the ground truth on the coronal plane and the sagittal plane are $0.487°$ and $0.471°$, respectively, which are acceptable in clinical applications.

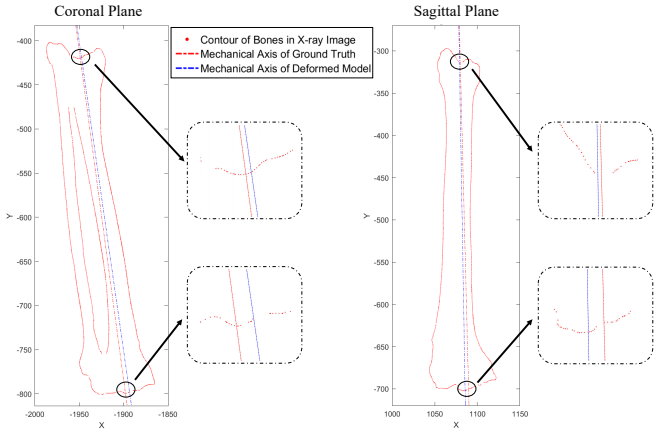


Fig. 5: Mechanical axes in coronal plane and sagittal plane.

C. Cadaver Experiment

Two different views of X-ray images were captured from a cadaveric tibia (with fibula). The general model used is the same as that in the simulation. Before the deformation, we extracted the intrinsic camera parameters, the bone contour observations in the X-ray images, camera poses, and aligned the general model into the CT (used as the ground truth for comparison) coordinate. A digital radiography machine using perspective projection took the X-ray images from two different viewing angles, and then their contours were manually extracted. Since the original cadaver CT has a connection with other objects, Meshlab [18] has been used for segmentation to obtain the ground truth model.

Because the X-ray images were received in DICOM (Digital Imaging and Communications in Medicine) format, the intrinsic camera parameters of the two images can be directly obtained. Besides, the approximate initial camera poses were also obtained from DICOM file. In addition, the general model has been transferred to the ground truth coordinate by performing the ICP algorithm [15].

Our proposed method was compared to the back-projection (BP) method [6] and method in [13] to reconstruct models on the Cadaver dataset. The mean reconstruction errors are $1.65mm$, $8.68mm$ and $3.78mm$ for the three methods, respectively. The estimated mechanical axes errors will be unacceptable if the reconstructed models have large errors. Thus, we only calculate the mechanical axes for our proposed method and the angles intersected by the estimated mechanical axis and the ground truth one are $0.758°$ and $0.562°$ in coronal and sagittal plane, respectively.

D. In-vivo Experiments

In-vivo datasets obtained from four patients are used to test the potential clinical value of the proposed framework, and each in-vivo dataset contains two real X-ray images and

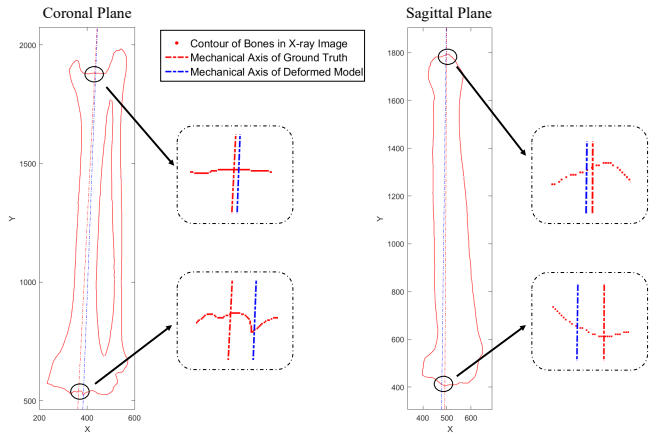


Fig. 6: Mechanical axes of one in-vivo dataset in coronal plane and sagittal plane.

the lower leg CT scans used as ground truth for comparison, where the X-ray images were taken by a digital radiography machine (orthogonal projection).

TABLE II: Summary of the mean reconstruction errors in in-vivo experiments.

	BP Method	Method [13]	Our Method
Patient 1 (mm)	6.193	1.989	**1.285**
Patient 2 (mm)	3.608	2.545	**1.220**
Patient 3 (mm)	5.274	2.399	**1.107**
Patient 4 (mm)	2.483	2.294	**1.024**
Average (mm)	4.390	2.307	**1.159**

Table II summarises the mean reconstruction errors from our method, the BP method used in [6] and the method in [13], and our method shows the highest reconstruction accuracy. Table III summarises the errors of estimated mechanical axes using our method. The mechanical axes for one in-vivo dataset on coronal plane and sagittal plane are described by Fig. 6, which indicates a minor deviation between the constructed model and ground truth.

TABLE III: Summary of the mean errors of estimated mechanical axes using our proposed method in in-vivo datasets.

Patient	#1	#2	#3	#4	Average
Coronal Plane (°)	0.596	1.250	0.265	0.704	0.704
Sagittal Plane (°)	0.483	0.151	0.717	0.424	0.444

The SSM-based and SSIM-based methods were not compared in this research because of the significant computational cost and insufficient models. However, the computing time used by BP method, method in [13], and our method were recorded and compared in Table IV. The time took by BP method was significantly higher than the method in [13] and our method because BP method solved the parameters by numerical optimisation. The time took by method in [13] was slightly higher than our method because it took more steps to reach an algorithm convergence. The reconstructed bones from the three methods were showed in Fig. 7, which indicated that the BP method and the method in [13] did not deform the top and bottom parts in a good quality, but the reconstructed model with our method was almost overlapped with the ground truth.

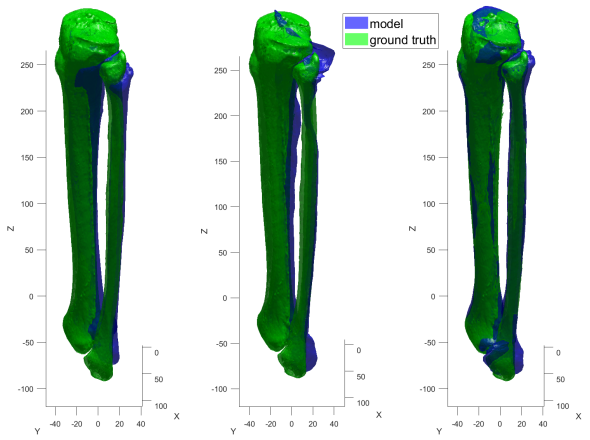


Fig. 7: Comparison of the ground truth (green) with the reconstructed models by the BP method (left), the method in [13] (middle), and our method (right) in in-vivo experiment.

TABLE IV: Computational time of different methods (in-vivo experiments)

	BP Method	Method [13]	Our Method
Patient 1 (s)	62,037	332	**140**
Patient 2 (s)	46,259	402	**381**
Patient 3 (s)	44,654	301	**206**
Patient 4 (s)	69,146	379	**252**
Average (s)	55,524	354	**245**

V. CONCLUSIONS

This paper proposes a framework which can reconstruct a patient specific 3D tibia (with fibula) model using a general 3D model and two X-ray images captured from the patient's lower leg. It uses deformation graph and estimates the deformation parameters to deform the general model to the patient specific model. The algorithm is compatible to both perspective and orthogonal projection models. Simulations results demonstrated the robustness and accuracy of the proposed framework. The cadaver experiment and in-vivo experiments showed the effectiveness and potential clinical value of the proposed framework.

The proposed framework can also be used to reconstruct other bones, not only tibia and fibula. We are planning to perform more evaluations on those in the near future. The proposed algorithm also has some limitations. For example, it requires the knowledge of the X-ray poses and that the general 3D model should have a similar shape as the one to be reconstructed. In the future, we are planning to use multiple 3D models and let the algorithm learn to use the most appropriate one to reconstruct the patient's 3D model in the system. In that way, the quality of the reconstruction could be further improved.

REFERENCES

[1] J. T. Evans, R. W. Walker, J. P. Evans, A. W. Blom, A. Sayers, and M. R. Whitehouse, "How long does a knee replacement last? a systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up," *The Lancet*, vol. 393, no. 10172, pp. 655–663, 2019.

[2] M. B. Santoso and L. Wu, "Unicompartmental knee arthroplasty, is it superior to high tibial osteotomy in treating unicompartmental osteoarthritis? a meta-analysis and systemic review," *Journal of orthopaedic surgery and research*, vol. 12, no. 1, pp. 1–10, 2017.

[3] B. Feng, W. Zhu, Y.-Y. Bian, X. Chang, K.-Y. Cheng, and X.-S. Weng, "China artificial joint annual data report," *Chinese medical journal*, vol. 134, no. 06, pp. 752–753, 2021.

[4] J. Gao, D. Xing, S. Dong, and J. Lin, "The primary total knee arthroplasty: A global analysis," *Journal of orthopaedic surgery and research*, vol. 15, no. 1, pp. 1–12, 2020.

[5] Y. Y. Kim, H. J. Shin, M.-J. Kim, and M.-J. Lee, "Comparison of effective radiation doses from x-ray, ct, and pet/ct in pediatric patients with neuroblastoma using a dose monitoring program," *Diagnostic and Interventional Radiology*, vol. 22, no. 4, p. 390, 2016.

[6] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, and M. A. G. Ballester, "A 2d/3d correspondence building method for reconstruction of a patient-specific 3d bone surface model using point distribution models and calibrated x-ray images," *Medical image analysis*, vol. 13, no. 6, pp. 883–899, 2009.

[7] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. Giphart, W. J. Niessen, and B. P. Lelieveldt, "2d–3d shape reconstruction of the distal femur from stereo x-ray imaging using statistical shape models," *Medical image analysis*, vol. 15, no. 6, pp. 840–850, 2011.

[8] G. Zheng, "Personalized x-ray reconstruction of the proximal femur via intensity-based non-rigid 2d-3d registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2011, pp. 598–606.

[9] C. J. F. Reyneke, M. Lüthi, V. Burdin, T. S. Douglas, T. Vetter, and T. E. Mutsvangwa, "Review of 2-d/3-d reconstruction using statistical shape and intensity models and x-ray image synthesis: Toward a unified framework," *IEEE reviews in biomedical engineering*, vol. 12, pp. 269–286, 2018.

[10] H. Kim, K. Lee, D. Lee, and N. Baek, "3d reconstruction of leg bones from x-ray images using cnn-based feature analysis," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2019, pp. 669–672.

[11] Y. Kasten, D. Doktofsky, and I. Kovler, "End-to-end convolutional neural network for 3d reconstruction of knee bones from bi-planar x-ray images," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2020, pp. 123–133.

[12] R. Shiode, M. Kabashima, Y. Hiasa, K. Oka, T. Murase, Y. Sato, and Y. Otake, "2d–3d reconstruction of distal forearm bone from actual x-ray images of the wrist using convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

[13] Y. Zhang, L. Zhao, and S. Huang, "Aortic 3d deformation reconstruction using 2d x-ray fluoroscopy and 3d pre-operative data for endovascular interventions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2393–2399.

[14] Y. Zhang, R. Falque, L. Zhao, S. Huang, and B. Hu, "Deep learning assisted automatic intra-operative 3d aortic deformation reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 660–669.

[15] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[16] R. W. Sumner, J. Schmid, and M. Pauly, "Embedded deformation for shape manipulation," in *ACM siggraph 2007 papers*, 2007, pp. 80–es.

[17] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on information theory*, vol. 29, no. 4, pp. 551–559, 1983.

[18] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia *et al.*, "Meshlab: an open-source mesh processing tool." in *Eurographics Italian chapter conference*, vol. 2008. Salerno, Italy, 2008, pp. 129–136.