Follow the Curve: Robotic-Ultrasound Navigation with Learning Based Localization of Spinous Processes for Scoliosis Assessment

Maria Victorova, Michael Ka-Shing Lee, David Navarro-Alarcon and Yongping Zheng

Abstract—The scoliosis progression in adolescents requires close monitoring to timely take treatment measures. Ultrasound imaging is a radiation-free and low-cost alternative in scoliosis assessment to X-rays, which are typically used in clinical practice. However, ultrasound images are prone to speckle noises, making it challenging for sonographers to detect bony features and follow the spine’s curvature. This paper introduces a robotic-ultrasound approach for spinal curvature tracking and automatic navigation. A fully connected network with deconvolutional heads is developed to locate the spinous process efficiently with real-time ultrasound images. We use this machine learning-based method to guide the motion of the robot-held ultrasound probe and follow the spinal curvature while capturing ultrasound images and correspondent position. We developed a new force-driven controller that automatically adjusts the probe’s pose relative to the skin surface to ensure a good acoustic coupling between the probe and skin. After the scanning, the acquired data is used to reconstruct the coronal spinal image, where the deformity of the scoliosis spine can be assessed and measured. To evaluate the performance of our methodology, we conducted an experimental study with human subjects where the deviations from the image center during the robotized procedure are compared to that obtained from manual scanning. The angles of spinal deformity measured on spinal reconstruction images were similar for both methods, implying that they equally reflect human anatomy.

I. INTRODUCTION

While ultrasound (US) has been proven to be a safe and reliable technique for scoliosis assessment [1], it is difficult for operators to identify anatomical features in ultrasound images due to its inherent speckle noises. The correct detection of bone features is crucial to properly follow the spinal curvature of scoliosis patients during the scanning. Thus, the quality of the resulting 3D spinal reconstruction is highly dependant on the sonographer’s experience.

This limitation can be overcome by using computer control to locate these features and guide a robotic arm that manipulates an ultrasound probe. To ensure that the spine is always in the probe’s field of view, the robot needs to follow the spine’s profile during scanning so that the vertebrae are located in the center of the ultrasound image. A spinous process is located in the middle of the vertebral body and indicates the vertebra’s presence in the image, whereas its absence indicates an intervertebral gap (see Fig. 1). These features are utilized for 3D spine reconstruction and scoliosis assessment by measuring the angle of the spinal curvature.

There is a number of researchers that have used Phase Symmetry (PS) as a set of filters to enhance bony structures, which can be used in image processing techniques to locate spinal features [2], [3]. Other work [4] has combined PS with machine learning (e.g., linear discriminant analysis classifier) to segment anatomical features in a US image frame as spinous processes, acoustic shadows, and other tissues. Another method [5] have used fully convolutional network (FCN) together with phase symmetry to segment the bone surface in a US image. The input image to this FCN has three channels, where the first channel is an original image, the second is the PS-processed image, and the third is the image resulting from computing the confidence map [6]. Methods based on pure deep learning techniques are few and are mainly focused on detecting the spine in different probe orientations (e.g., sagittal). These are typically used for spinal injections where the features appear different and are not suitable for scoliosis assessment [7], [8].

There are various limitations with the above state-of-the-art methods, e.g., the considerable uncertainty in the estimation where is the intervertebral gap; the need to segment pixels containing bony structures manually; multiple parameters that must be tuned in all three different algorithms (i.e., PS, confidence map, and FCN). Therefore, these methods are not suitable for continuous spinal scanning.

There is currently no algorithm designed explicitly for spinous process localization (representing the center of the vertebrae) that we can use to track the spine in our scoliosis application. Therefore, we take inspiration from the area of human pose estimation, which also requires accurate landmarks localization in the form of joints. One method frequently used in this field is landmark pose regression from heatmaps generated by FCN, where each pixel represents a probability of containing the sought landmark. Various researchers have adopted this approach for localization of these types of landmarks [9]–[11].

The original contribution of this study is a spinal curvature following approach with intelligent robot-driven ultrasound for scoliosis assessment. Our developed prototype is guided by real-time ultrasound images and has spinal curvature following capabilities. To accurately detect bone features from ultrasound images, our method uses a robust deep learning based network specifically designed for spinous process localization. The network takes as input raw ultrasound images, and it outputs a spatial heatmap that indicates the location probability of the spinous process; The maximum...
probability predicts the position of these bony features on each ultrasound image. The predicted location enables the system to guide the robotic probe along the curved spine by using an automatic motion controller combined with a Kalman filter estimator, which forms a continuous spinal path compensating for the intervertebral gaps. The proposed robotic scanning approach was evaluated on both spinal phantoms and human volunteers; its performance was compared with the conventional manual scanning approach.

The rest of this manuscript is organized as follows: Sec. II presents the methodology. Sec. III reports the conducted experiments. Sec. IV gives final conclusions.

II. METHODS

A. Robotic-ultrasound scoliosis examination procedure

During the regular ultrasound scoliosis assessment developed by Zheng et al., [1] sonographer manually scans the human spine with an ultrasound probe in caudo-cranial direction, trying to center the spine in the field of view of the probe. The sonographer has to apply pressure during the scan and rotate it to ensure that the probe maintains tight contact with the patient’s skin. The ultrasound system used for manual scanning is Scolioscan Air [12], which consists of a USB Ultrasound probe and a tablet with Scolioscan Air software, which receives images and coordinates of the probe for spinal 3D reconstruction.

The robotic scanning process, (the setup is shown in Fig. 2 b)), begins by manually positioning the robotic arm operated ultrasound probe at the sacrum level; from there, the robotic arm travels towards the subject’s back until the force meets the force setpoint. The probe then begins to go higher along the subject’s back. The robot provides continual pressure to the subject’s back during the motion and follows the spinous process as a reference characteristic of the spine’s center. The robotic procedure uses a pitch rotation to keep the probe normal to the subject’s back and a yaw rotation to make sure the left and right sides of the probe surface are in contact with the skin. After both manual and robotic procedures, the software generates coronal images as slices of the 3D spinal reconstruction for curvature measurement, according to the spinous process angle method presented in [1]. The angle is calculated between the most tilted vertebrae of the scoliotic curve. During both scanning approaches, human subjects could breathe normally, and the posture was fixed by asking the subject to lean on the built-in supporters as in Fig. 2 b).

B. Materials and experimental setup

Fig. 2 b) shows the setup for automatic scoliosis assessment using a robotic arm - (Universal robot UR5) with force sensor mounted (FT300, Robotiq) and USB ultrasound probe (Sonoptek, Beijing). The Ultrasound probe (Fig. 3 b)) captures images with 7.5 MHz, depth of 6 cm, and sends raw data at 10 fps over USB to PC, where the images are formed in a size of 640x480 pixels. The probe’s aperture is rectangular with a more extended size of 80mm. The robot was connected via TCP/IP protocol with a PC, where the control algorithm was launched at a rate of 30 fps Hz. Since the robotic control has a higher frequency than the ultrasound images frame rate, several robotic positions correspond to a single image.

C. Coordinate systems

Fig. 3 a) shows different coordinate systems for robotic-US setup. The Base coordinate system is fixed for all times with the origin at the point where the robot is attached to the working table. The Init coordinate system is initialized upon the start of each scanning procedure, placing the US probe perpendicular to the subject’s back and parallel to the ground to match the settings of manual scanning. Tool Central Point (TCP) coordinate frame coincides with the Init frame upon start, but as the US probe moves along the subject’s back, the TCP frame moves together with the US probe. The TCP frame is used for control purposes. In Fig. 3 b) the rotations of the probe in TCP coordinate frame are presented. The ultrasound images have two axes, X is along the image width, and Y is along the height, and, at the same time, it

---

1Ethical approval HSEARS20210417002 was given by Departmental Research Committee (on behalf of PolyU Institutional Review Board)
represents the penetration depth in spinal tissues. The image has a resolution of 640x480 pixels.

D. Dataset

The acquired data set consists of 24 scoliosis patients, which were scanned manually by a sonographer. Each ultrasound frame for each subject was manually labeled by an ultrasound expert indicating the spinous process location, if present, as shown in Fig. 1 c). The target heatmap images were generated by applying a 2D Gaussian distribution aligned with the manual label’s center. The resulting dataset consists of spinous process images (not including intervertebral images, where the spinous process is absent) and correspondent target heatmap images. The training subset has 19 subjects, 9,396 spinous process images, and the test has five subjects of 2,972 spinous process images.

E. Spinal features localization

The method of spinous process localization uses the heatmap approach. The heatmap represents the probability of each pixel of the image to be a spinous process, forming the Gaussian distribution around the point with the maximum probability; that point is the sought spinous process location. The schematic overview of our proposed network for spinous process classification and localization is conceptually shown in Fig. 4. The widely used ResNet is used as a backbone to extract image features from input ultrasound images. The conv1 to conv5 layers in the figure represent the five convolutional stages in ResNet.

We replaced the last fully-connected layer of the ResNet (which gives the class distribution score) with three deconvolutional layers with batch normalization [13] and ReLU activation [14], which act as the decoder to generate the image features heatmap [10]. Each deconvolutional layer has 256 filters with a 4x4 kernel; the stride is 2. The deconvolutional layers are followed by one 1x1 convolutional layer, which transforms the resulted features matrix to a final heatmap where each image pixel intensity represents the probability of being one of the classes; in our case, the maximum intensity represents the high probability of the pixel belonging to a spinous process class. The loss between the ground truth and predicted heatmap is calculated using the mean squared error (MSE). The final location is derived from the maximum intensity pixels of the resulted heatmap. To increase the robustness of the dataset, we use various image data augmentation techniques, such as rotation, horizontal and vertical flipping.

F. Spinal features tracking

The sequence of US images for one human subject is called ultrasound sweep. In the sweep, the spinal images of the spinous process alternate with intervertebral gap images (shown in Fig. 5). Since the localization network does not provide meaningful location information for intervertebral gap (the detection probabilities are less than 50% and thus
the locations are rejected), there is a need for a method that can form a continuous path. To overcome the issue with missing location information, we generate Kalman computed continuous spinal path. Kalman filter [15] was used to fuse filter-predicted locations \( \hat{x}_{k-1} \) with network detected locations \( x_{net} \), whenever they are available, as expressed in eq. 1. Here \( K_k \) is Kalman gain, and \( \hat{x}_k \) is a resulted next-frame location. This method also helps to filter the network output to generate smoother locations as inputs for image-based robotic control.

\[
\hat{x}_k = \hat{x}_{k-1} + K_k (x_{net} - \hat{x}_{k-1}) \tag{1}
\]

During the ideal scanning procedure, the spinal path plot will be a straight line positioned at the center of the coordinates (half of the image width size), which corresponds to the scanning when the spinous process was kept at the center of the field of view of the ultrasound probe. Any deviations of the points in the plot from the center of the coordinates can be considered inaccuracies of the scanning procedure.

### G. Learning based robotic navigation

Ultrasound feature-based navigation control works along X-axis of TCP coordinate frame \( v_x \). Each captured ultrasound frame used as an input for pre-trained deep learning model (Fig. 4), which outputs the predicted location of the spinous process and probability of this point belonging to the spinous process (confidence of prediction). The locations with probability higher than the threshold considered to be valid detected spinous process. Since the resulted heatmap is 56x56, the location output should be re-scaled to correspond to original image size of \( w_{image} = 640 \), \( h_{image} = 480 \). Those locations are fused with Kalman filter predicted locations generating the location coordinate \( x_{image} \) for robotic control, expressed in coordinate frames as shown in Fig. 3 c). The error between the spinous location and the image center \( \delta x_{pixels} = x_{image} - w_{image}/2 \), can be re-scaled to be expressed in meters, which is the same dimensions as the output velocity \( \delta x = \frac{\delta x_{pixels} \cdot \delta w_{probe}}{w_{image}} \). The proportional control law using the spinous process location as feedback would look as following: \( v_x = -K_{im} \delta x \), where \( K_{im} \) is a proportional control gain. There are two different control gains used depending on the distance to the desired spinous location. This approach is made to prevent sudden velocities generated in reaction to far located points. Adding exponential smoothing with control gain \( \alpha \) to the control law to ensure smoother velocity output would transform this equation to following:

\[
v_x(t) = \alpha(-K_{im} \delta x) + (1 - \alpha)v_x(t - 1) \tag{2}
\]

### H. Force-driven acoustic coupling

The manipulation of US probes over the body involves two main problems: (1) stable interaction with deformable tissues; and (2) feature-based navigation control. The former is needed to ensure that the probe provides stable contact with human skin surface to get the US images of acceptable quality; The latter is needed to autonomously conduct the scanning procedure following the spinal curvature. Thus, a velocity control during the US-guided procedure uses mainly two parameters to form a control loop: force screw \( f = (f_x f_y f_z m_x m_y m_z)^T \) and spinous process location in each US images \( X_{im} \). The control velocity vector is expressed as \( v_{TCP} = (v_x v_y v_z r_x r_y r_z)^T \). During the scanning, the robot moves along the spine in the direction on the Y-axis of TCP coordinate frame with constant velocity \( v_y \) set before scanning. PID velocity control is used for force control along Z-axis of TCP coordinate frame \( v_z \) according to the Fig. 3 similar to [16]. While the traditional PID control works well for the applications where setpoint (in our case it is \( F_{ref} \)) is not changing, however when the setpoint changes, the sudden spike appears in the velocity output. The setpoint change necessity comes from an idea of changing force settings according to the spinal region (see section II-J). One way to overcome this issue is to use “derivative on measurement”, which neglects the reference force change, \( \frac{dF_{ref}}{dt} = 0 \). Since \( \frac{df}{dt} = \frac{dF_{curr}}{dt} - \frac{dF_{curr}}{dt} \), the resulted equation:

\[
v_z = K_p e + K_i \int e dt - K_d \frac{dF_{curr}}{dt}, \tag{3}
\]

where \( e = F_{ref} - F_{curr} \) is an error between reference force \( F_{ref} \), set prior to scanning, and current force measurement \( F_{curr} = f_z \), coming from force sensor in the Z direction. The control gains \( K_p, K_i, K_d \) are chosen experimentally according to the intrinsic parameters, such as the robotic system response and extrinsic parameters, such as stiffness of the subject’s back.

In addition to a PID control, the saturation \( v_{lim} \) of the output is used as a safety measure to prevent larger velocities from affecting the stable contact with the skin surface or causing any discomfort for the subject during the scan. There is also a custom set limit for the maximum force exerted on the robot end-effector. If \( F_{crit} \) is reached, the robot stops the operation to ensure the safety of the patient and operator.

### I. Force-based probe orientation adjustments

To ensure the robotic actuated ultrasound probe smoothly adapts to the curvature change of the human back surface, the moment (torque) \( m_z \) induced control is used. The example of the procedure with the force-compliant scanning is presented in Fig. 6. The coordinate frames of the force sensor and TCP frame used to control coincide; thus, the \( r_x \) rotation will affect readings of \( m_z \). The \( r_x \) rotation change during the scanning ensures that the ultrasound probe faces the human back surface at an angle close to 90°. The proposed control follows the following law: \( r_x = -K_{pitch} m_z \).

The principle behind this law is that the rotation \( r_x \) will compensate the moment \( m_z \) change during the scanning. The control gain \( K_{pitch} \) specifies how sensitive it would be the rotation and how fast the compensation would occur.

### J. Spinal regions classification for scanning settings adjustments

Due to anatomical spine structure, the lumbar region of the spine usually lays deeper under the layer of the fat and muscles than the thoracic region of the spine. According to these anatomical differences, the pressure applied by the
sonographer during the scanning to those regions is slightly different; the lumbar region of the spine requires higher forces applied to compress the layer of fat and capture the entire bony structures. We propose using the network for spinal regions classification to detect when the lumbar region ends and thoracic starts. The robotic scanning settings such as force applied and ultrasound settings as ultrasound gain could be adjusted by detecting this region change.

The proposed network has two heads for two different tasks, classification and localization; the first is a deconvolutional head for localization task is similar to one presented in Fig. 4; the second is a classification head used to output the probability of an image to belong to a certain spinal region. The diagram of resulted multi-task network is presented in Fig. 7. The network is trained end-to-end, which means the total loss combined from two network heads is used for back-propagation. The total loss is a sum of two losses with a scaling factor, $C$, which is used to balance the magnitudes of the two losses, $TotalLoss = Loss_{classification} + C \times Loss_{localization}$. The resulted class number where the image frame belongs can be found by applying a softmax layer on the probabilities output from the classification head. The loss for localization task is MSE and for classification is Cross-entropy loss.

III. RESULTS

A. Spinal features localization method’s results

The model presented in Fig. 4 was trained for 100 epochs reaching the best validation accuracy at the 10th epoch; learning rate was 0.0001, batch size is 12. Adam [17] optimizer was used. The initial weights for ResNet [18] backbone were loaded from publicly available ImageNet pre-trained ResNet18 model. The inputs for the model were ultrasound images of the spinous process collected from 19 subjects. The images were normalized (so that the pixels becoming floats in range $[0, 1]$) and resized to a size of 224x224 pixels. The targets were heatmaps generated from manually labelled images, as shown in Fig. 9 a) and Fig. 9 b). The resulted model was tested on spinous process images collected from 5 subjects (2971 images).

The accuracy of detected spinous process locations is calculated similarly to a PCK (percentage of correct key points) metric used in human pose estimation tasks. The detected spinous process is considered correct if the distance between the predicted and target spinous process is within a threshold, here 50%. The resulting accuracy on a test set (2,972 SP images) was 97.8% with a mean distance error of 8 pixels, corresponding to 1.0mm.

B. Spinal features tracking method’s results

Using the obtained model on a sequence of ultrasound images obtained from one human subject (referred to as ultrasound sweep) results in a sequence of points of predicted locations and the corresponding probability, which expresses the confidence of predicted location. The Fig. 10 b) LEFT demonstrates the spinous process presence in ultrasound sweep, obtained by manual scan. Red dots correspond to the predicted locations with the confidence of more than 50%;
phantom robotic scanning with speed of $v_{lim} = 0.05$, Force $F_{ref} = 7N$, PID control $K_p = 0.0004$ and tilt coefficient $K_{pitch} = 0.07$. Kalman filter parameters were $Q = 0.5, R = 500$. The phantom scan: spinous process location mean $x_{image} = 322px$, with $STD = 52px$ (6.5mm).

Fig. 12 shows the results of robotic scanning for human subject with mild severity of scoliosis (i.e. the deformity angle is less than 25 degrees). The plots display the Kalman computed paths for both robotic and manual scanings. The path for the manual scanning was computed after the manual procedure was performed, while for robotic approach the path was used for the navigation purpose. The scanning settings were the following: speed of $v = 0.004$, velocity output clipping of $v_{lim} = 0.002$, force setpoint $F = 15N$ (the force range was chosen according to our previous work on spinal scanning [19]), PID control $K_p = 0.0003, K_d = K_i = 0.00003$ and tilt coefficient $K_{pitch} = 0.07$. Kalman filter parameters for spinous process location output were $Q = 0.5, R = 500$. A mean deviation from the image center for the resulted spinous process path was calculated as $\delta x_{mean} = \frac{1}{N} \sum(x_i - x_{image})/2$. For this case the $\delta x_{mean} = 7.8px$ (1.0mm) with $STD = 6.5px$ for robotic approach and $\delta x_{mean} = 36.8px$ (4.6mm) with $STD = 36.9px$ for manual scanning.

The other example of the human subject with moderate scoliosis (i.e. the deformity angle is more than 25 degrees) is displayed at Fig. 13. For this case the $\delta x_{mean} = 22.6px$ (2.8mm) with $STD = 14.7px$ for robotic approach and $\delta x_{mean} = 34.7px$ (4.3mm) with $STD = 31.5px$ for manual scanning.

D. Spinal regions classification for parameters adjustments

For region detection network was trained to classify ultrasound images of the ultrasound scan into four classes: Intervertebral gap, Lumbar spinous process, Sacrum, Thoracic spinous process. Together with class prediction, the network outputs the heatmap, where the maximum intensity corresponds to the spinous process location with a certain probability (the magnitude of the maximum intensity). The network was trained on the dataset, consisting of the spinous
loss was of 0.5 every 20 epochs. The weight loss for the heatmap 12, Adam [17] optimizer was used with learning rate decay was achieved on 60 epoch with projects (18,939 images) were used for training. The best model process, sacrum, gap ultrasound images. A total of 15 sub- b) manual scanning.

of the human's spine from images collected during a) robotic scanning and the green dash line is a middle line of the image. Resulted reconstruction Kalman computed path (in pixels) frame by frame during the scan, where Fig. 12: Spinous process tracking results on human with mild Scoliosis. the result of phantom scanning shows that the mean spinous process location among ultrasound frames was in $x_{image} = 322$ pixels, which is similar to the center of the ultrasound image $x_{image} = 320$ pixels. Visually assessing the tracking results for phantom, in Fig. 11, it is clear that the coronal image generated from the robotic scan has spine centered in the field of view of the probe, the spinal features are more prominent, and the edges of the image are smoother. Since phantom is an idealistic case of the spine, the navigation performance was further assessed on human subjects.

The human tracking results show how well the robotic approach could maintain the spinous process in the middle of US frames for participants with mild and moderate scoliosis (Fig. 12 and Fig. 13). The mean deviation from the image center of the detected spinous process during the scanning was between 1.0 mm for mild and 2.8 mm for the moderate case. The more significant deviation for the moderate case is explained by the greater angle of spinal deformity, where the robot takes longer time to center the spinous process in the field of view. Visual assessment of the image quality by the sonographer experts states that the robotic image for both cases looks smoother, the spinal features are more distinguishable. The last comment can be explained by the optimal force distribution along the spine. The robot applies a constant force, which is also generally slightly higher than what human does.

The region pitch rotation adjustment turned to be an essential improvement for the system. Due to the probe’s shape, the higher rotations cannot be used if the subject has significant lordosis at the lumbar region. Then lower rotation gain should be used $r_x$. However, if the same subject has a highly curved upper back, the lower rotations will not be sufficient, and the probe will lose contact with the subject’s back. Thus, region-specific pitch rotation can significantly improve the quality of resulted scanning images.

The spinal region’s network with or without “intervertebral gaps” both perform well according to accuracy; the network without gaps was chosen since it has an output more straightforward for the lumbar and thoracic region splitting. The robotic experiment with region-specific pitch rotation clearly shows the improvement in the resulted coronal image, Fig. 14. The robot could smoothly follow the spinal curvature, maintaining tight contact with the subject’s skin, avoiding the air gap in the spine’s thoracic (upper) region.

There are several limitations of the current work. There was no external visual feedback, which could be helpful...
Fig. 14: Example scan of human subject. Spinal regions network output for 2-head network with classification head of 3 classes. The resulted spinal coronal images with fixed rotation (LEFT) and with region specific rotation (RIGHT).

When the subject has a narrow space between back bones (scapula) and the additional roll rotation is needed. Currently, there is no roll rotation designed in this study, which could allow aligning probe with the direction of the SP path. There were no severe scoliosis patients involved in this study, where due to the extreme spinal deformity, the spinous process is not present in the transverse ultrasound images for some vertebrae. The project’s next step is in progress: the system validation and reliability assessment in the subject’s trials. It is vital to determine user cases where robotic scanning is practical or only the manual can be used, opening the path for future system improvements.

REFERENCES


