IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 24, 2022, accepted April 5, 2022, date of publication April 8, 2022, date of current version April 19, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165936

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

MARIA VICTOROVA[1], MICHAEL KA-SHING LEE[1], DAVID NAVARRO-ALARCON[1,2], (Senior Member, IEEE), AND YONGPING ZHENG[1], (Senior Member, IEEE)

[1]Department of Biomedical Engineering, The Hong Kong Polytechnic University (PolyU), Hong Kong
[2]Department of Mechanical Engineering, Research Institute for Smart Ageing, The Hong Kong Polytechnic University (PolyU), Hong Kong

Corresponding author: Maria Victorova (maria.victorova@connect.polyu.hk)

This work was supported in part by the Research Impact Fund (RIF) of the Hong Kong Research Grant Council under Grant R5017-18F, and in part by The Hong Kong Polytechnic University through the Intra-Faculty Interdisciplinary Project under Grant ZVVR.
This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethical approval HSEARS20210417002 was given by the Departmental Research Committee on behalf of PolyU Institutional Review Board.

ABSTRACT The scoliosis progression in adolescents requires close monitoring to timely take treatment measures. Ultrasound imaging is a radiation-free alternative in scoliosis assessment to X-ray, which is typically used in clinical practice. However, ultrasound images are prone to speckle noise, making it challenging for sonographers to detect bony features and follow the spinal curvature. This study introduces a novel robotic ultrasound approach for spinous process localization and automatic spinal curvature tracking for scoliosis assessment. The positions of the spinous processes are computed using a fully connected network with a deconvolutional head. A 5-fold cross-validation was performed on a dataset of ultrasound images from 25 human subjects with scoliosis. The resulting percentage of correct keypoints of the spinous process is 0.966 ± 0.027 with a mean distance error of 1.0 ± 0.99mm. We use this machine learning-based method to guide the motion of the robot-held ultrasound probe and to follow the spinal curvature while capturing ultrasound images. We present a new force-driven controller that automatically adjusts the pose and orientation of the probe relative to the skin surface, which ensures a good acoustic coupling between the probe and skin. We extended the network architecture to additionally perform classification of the spine into its regions, i.e., sacrum, lumbar, and thoracic, which are used to adjust the probe's orientation to account for the varying curvature along the spine. After the autonomous scanning, the acquired data is used to reconstruct the coronal spinal image, where the deformity of the scoliosis spine can be assessed and measured. The proposed learning-based method for anatomical landmarks localization was compared to conventional methods based on phase symmetry and image intensity. The learning-based method proved to be more precise for spinous process localization while processing images at a faster rate, which is advantageous for real-time scoliosis scanning. To evaluate the performance of our robotic method, we conducted an experimental study with human scoliosis subjects where deviations of the spinous process from the image center can be compared to those appearing in a manual scan. Our results show that the robotic approach reduces the mean error of spinal curvature following for mild scoliosis from 4.6 ± 4.6mm (manual scanning) to 1.0 ± 0.8mm (robotic scanning); For moderate scoliosis from 4.3 ± 3.9mm (manual scanning) to 2.8 ± 1.8mm (robotic scanning). The angles of spinal deformity measured on spinal reconstruction images were similar for both methods, implying that they equally reflect human anatomy. The spinal region-specific moment-based probe orientation control showed to improve the scanning performance. An ablation study was performed to investigate the importance of each component of the proposed system.

INDEX TERMS Medical robots and systems, computer vision for medical robotics, ultrasound navigation, scoliosis, spinous process, robotic manipulation.

The associate editor coordinating the review of this manuscript and approving it for publication Junhua Li.

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 in ultrasound images during a transverse scan (probe oriented such that it divides the human body into upper and lower parts) 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 dependent on the sonographer's experience.

This limitation can be overcome by using computer algorithms to locate these features and guide a robotic arm manipulating the 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 (SP) is located in the middle of the vertebrae and indicates the vertebra's presence in the image, whereas its absence indicates an intervertebral gap (see Fig. 1). Therefore, these features can be used for navigation during scanning and subsequent 3D spine reconstruction, as well as assessment of scoliosis by measuring the angle of the spinal curvature.

There are 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]. Yu *et al.*, [4] used phase symmetry for US image preprocessing and template matching to extract features. The extracted features were then used for binary classification with a support vector machine to identify US frames suitable for epidural injection. Tran *et al.*, [5] also used the approach of PS and template matching for localization of injection point by identifying lamina in a sagittal view (probe oriented such that it divides the human body into left and right). Some other works have 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 [6]. This method required manual segmentation of each tissue in an ultrasound image. Another more recent method used a fully convolutional network (FCN) together with phase symmetry to segment the bone surface in US image [7]. The input image to this FCN had three channels, where the first channel was the original image, the second was the PS-processed image, and the third was the image resulting from computing the confidence map [8]. The authors discovered that the performance of the PS and confidence map methods varies greatly between different US machines. While the designed method outperformed the conventional method based only on PS, it took 7 seconds to process one frame (making it unsuitable for real-time navigation), and its reliance on initialization parameters for each individual sub-method was a limitation. The dataset was relatively small, consisting of around 1300 images. Methods based on pure deep learning techniques are few and are mainly focused on detecting the spine in sagittal probe orientation; These are typically used for spinal injections where the features appear differently and they are not suitable for scoliosis assessment [9], [10].

There are various limitations with the above state-of-the-art methods, e.g. the need to manually segment pixels containing bony structures, and the multiple parameters that must be tuned in all three different algorithms (i.e., PS, confidence map, and FCN), etc. Most crucially, as these approaches are based on the identification of spinous processes on static frames, they do not account for the continuous structure of the spine, where spinous processes alternate with intervertebral gaps.

According to the above-discussed methods for bony feature detection, there are no existing algorithms that are reliable and straightforward for spinous process localization using continuous US image streams along the spine. As a result, we take inspiration from recent work on human pose estimation, which requires the detection of precise landmarks (viz. human joints) [11]–[13]. This pose estimation method is based on a heatmap approach, in which the network generates an image in which each pixel represents the probability of belonging to a given class. The maximum intensity indicates the location of the sought-after landmark. For evaluation, the percentage of correct keypoints (PCK) [14] metric is used, which provides the percentage of detected landmarks that are within a distance threshold from the ground truth.

The most common approach for heatmap generation is the Hourglass model [11], which is a variation of a fully convolutional network. The peculiarity of the Hourglass model is that the decoding layers combine upsampling with the spatial information coming from the encoding layers; Thus, the resulting image has a higher specificity. The resulted total PCKh@0.5 (the threshold = 50% of the head size) was 90.9% on the MPII dataset [15]. This model performs well under a range of conditions, including high occlusion and several persons in close proximity. However, as a single landmark localization job, the detection of spinous processes will benefit from a lighter-model solution with greater simplicity. Thus, Xiao *et al.* [12] proposed a more optimal yet efficient model for landmark localization. It showed PCKh@0.5 of 92.3% on the MPII dataset.

Learning-based approaches have the potential to overcome the shortcomings of existing methods for fast detecting bony features, providing precise location of the spinous process itself rather than surrounding features; The methods include learnable parameters, which do not require hand-tuning of parameters. They also provide an accuracy value that can be used to determine whether an image belongs to a vertebra region or an intervertebral gap, which is critical for continuous spinal scanning.

A. OUR CONTRIBUTION

To the best of our knowledge, this is the *first study* that developed an automatic spinal curvature navigation system for scoliosis assessment using real-time ultrasound images. Compared to other methods for anatomical landmark

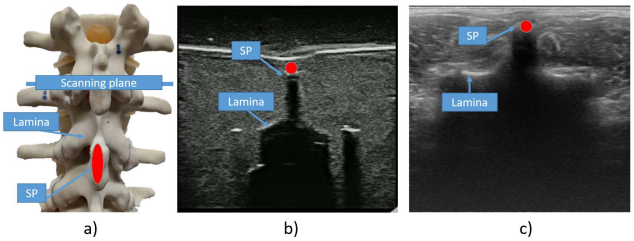


FIGURE 1. Spinal anatomical features on phantom and US images of phantom and human. Lamina and spinous process (SP). a) Spinal features presented at the spinal column phantom, b) Spinal features visible on US images of a phantom, c) Spinal features visible on US images of a human.

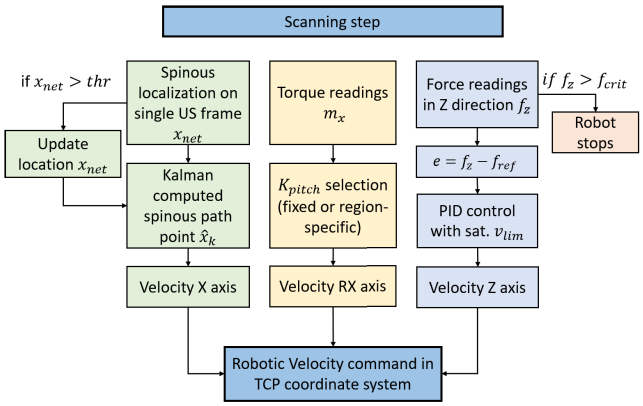


FIGURE 2. Overview of the proposed robotic scanning method to compute the robot velocity commands based on pose, force and ultrasound data.

localization, our new method uses deep learning to robustly localize landmarks, inspired by the work on human pose estimation [11]–[13], instead of the traditional phase symmetry or manual segmentation methods. We provide a comparison of the performance of those methods with our proposed method for spinous localization. Other methods do not address navigation, they detect features on static frames where the spinous process is present. However spinous process and intervertebral gap sectors alternate in the human spine, which complicates its localization based on US images. To deal with this issue, we propose a Kalman filter-based method that computes a continuous path and compensates for the gaps. Compared to other approaches [16]–[18], our method does not require any external vision sensors for trajectory pre-planning (which facilitates its implementation). In addition, the robot-guided ultrasound probe is able to maintain a stable acoustic coupling by pose and orientation adjustments based on contact force and moment feedback. To validate our method, we conducted a detailed experimental study with both spinal phantoms and human subjects. The performance of the system was compared to the conventional manual scanning approach. We also performed an ablation study to investigate the importance of each module of the system.

The rest of this manuscript is organized as follows: Sec. II presents the methodology (the overview is presented in Fig. 2. Sec. III reports the conducted experiments. Sec. IV gives conclusions.

II. METHODS

A. ROBOTIC-ULTRASOUND SCOLIOSIS EXAMINATION PROCEDURE

In state-of-the-art methods for ultrasound scoliosis assessment (see e.g., [1]), a sonographer manually scans the human spine with an ultrasound probe in a caudo-cranial direction, while centering the spine in the field of view of the probe. The sonographer has to apply pressure during the scanning and rotate it to ensure that the probe maintains tight contact (i.e., a good acoustic coupling) with the subject's skin and hence produce US images of acceptable quality.

The robotic scanning process (see Fig. 3b) begins by manually positioning the robot-held probe at the sacrum level. From there, the robotic arm travels towards the subject's back until a force setpoint is reached. The probe then begins to move upwards along the subject's back. The robot then provides continuous pressure during the motion and follows the spinal curve by tracking the spinous processes. The robotic procedure uses a "pitch" rotation R_x to keep the probe normal to the surface.

After both manual and robotic procedures, the software generates coronal images as slices of the reconstructed 3D spinal model [19], which are used for scoliosis assessment by measuring the curvature according to the spinous process angle method [1], [20]. This angle is calculated between the most tilted vertebrae below and above the scoliotic curve apex. During both scanning approaches, human subjects can breathe normally, and the posture is fixed by asking the subject to lean on built-in supporters, see Fig. 3b. For both scanning procedures, a water-based ultrasound gel was evenly applied to the subject's back with a sponge to act as a conductive medium for ultrasonic waves, as is a common practice for ultrasound examinations.

B. MATERIALS AND EXPERIMENTAL SETUP

Fig. 3b shows the setup for automatic scoliosis assessment using a robotic arm - (Universal robot UR5) with a force sensor (FT300, Robotiq) and USB ultrasound probe (Sonoptek, Beijing). The ultrasound probe (Fig. 4b) captures images at 7.5 MHz, with a depth of 6 cm, it sends raw data at 10 fps to a PC, where the images are formed in a size of 640 × 480 pixels. The probe's aperture is rectangular with a length of 80mm. The robot is connected via TCP/IP protocol with the PC, where the control algorithms are implemented at a rate of 30 Hz. The Scolioscan Air system that is used for manual scanning streams data at 60 fps, therefore, the robotic scanning system needs to operate twice slower than the manual scanning system to produce a similar number of frames per scan (the human subjects did not express any discomfort resulting from this longer robotic approach).

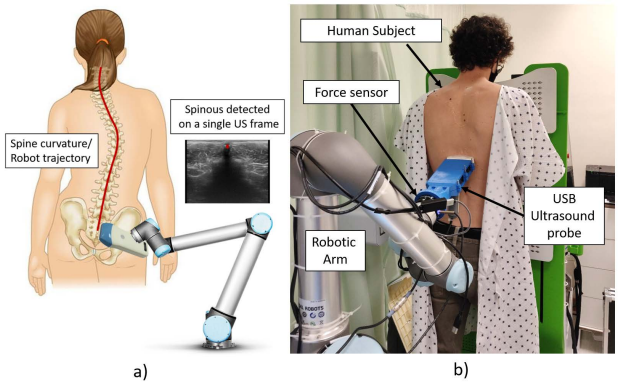


FIGURE 3. a) Proposed spinal curvature tracking. The robotic arm with ultrasound probe follows the spinal curvature by real-time detection of spinous processes on ultrasound frames. b) Setup for robotic scoliosis assessment for human scanning. Robotic arm UR5 presented with mounted force sensor FT300 and ultrasound probe in 3D printed probe holder. The human subject is stabilized by the supporters mounted on the frontal plate of the Scolioscan machine.

C. COORDINATE SYSTEMS

Fig. 4a shows different coordinate systems used for the robot. The "Base" coordinate system is fixed and represents the system's inertial frame. 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 the manual scanning. The "TCP" (tool central point) coordinate frame coincides with the Init frame upon start, and as the US probe moves along the back, the TCP frame moves with it. This TCP frame is used in our method for control purposes, where the force sensor measurements $f = [f_x, f_y, f_z, m_x, m_y, m_z]^T$ coincide with the directions of the TCP robot velocities $v = [v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^T$. In Fig. 4b, the rotations of the probe in the TCP coordinate frame are presented. The ultrasound images have two axes, X and Y, which represent the horizontal and vertical image coordinates, respectively; The Y coordinate also represents the penetration depth into the spinal tissues.

D. DATASET

The acquired dataset consists of 25 scoliosis patients,[1] which were scanned manually by a sonographer. The consent forms for all the participants were collected. There were 18 participants with mild scoliosis (angle of spinal deformity between 10° and 24° degrees) and 7 subjects with moderate scoliosis (angle between 25° and 37° degrees) in the dataset, the mean angle was 19.7 ± 5.7 degrees. Each ultrasound frame for each subject was manually labeled by an ultrasound expert indicating the spinous process location, if present, as shown in Fig. 1c. The target heatmap images were generated by applying a 2D Gaussian distribution aligned with the manual label's center. The resulting dataset consists of 425 spinous processes, 17 for each patient forming 13,674 spinous process images (not including intervertebral images, where the spinous process is absent) and their corresponding target heatmap images.

[1]Ethical approval HSEARS20210417002 was given by the Departmental Research Committee (on behalf of PolyU Institutional Review Board).

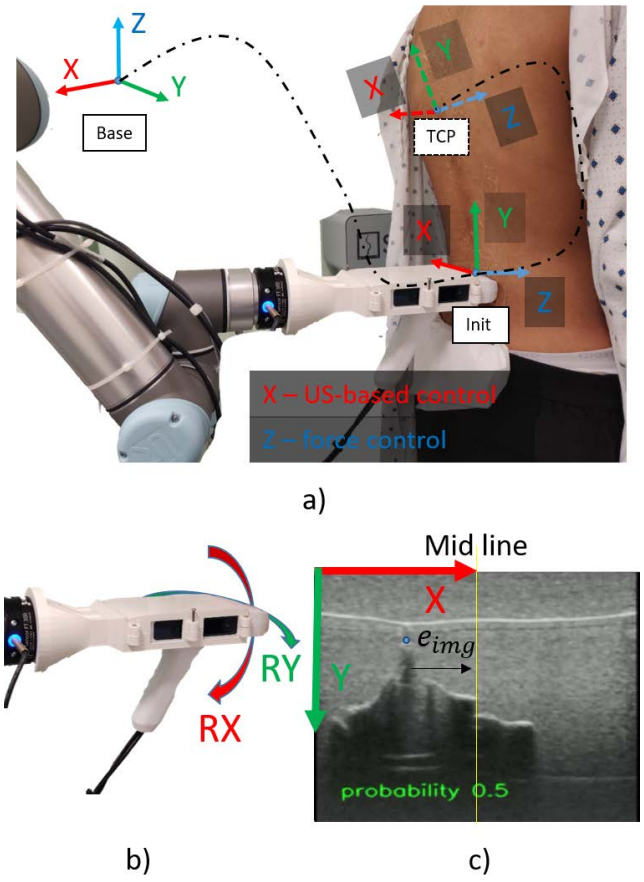


FIGURE 4. a) The robotic-US setup with coordinate frames labeled. b) Ultrasound Probe 3D-printed holder. c) B-mode US image coordinate frames.

E. LOCALIZATION OF SPINAL FEATURES

1) LEARNING-BASED METHOD

Our method developed for spinous process localization uses a heatmap approach and it is inspired by [12]. 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. 6. The widely used ResNet [21] 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. Other models, such as VGG11 and DenseNet121, were considered as backbones as well, but their performance on ultrasound images was found to be less efficient [22].

We replaced the last fully-connected layer of the ResNet (which gives the class distribution score) with three

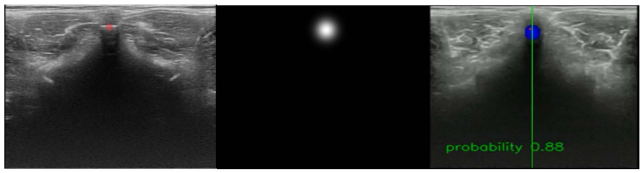


FIGURE 5. Results of the spinous process localization on a single frame. a) The US frame manually labelled; b) Ground truth mask: the resulted target heatmap with a Gaussian drawn around manual label center; c) localized spinous process with proposed architecture (red dot), ground truth Gaussian (blue) and middle line of the image (green). The resulted confidence of the localized point is 88%.

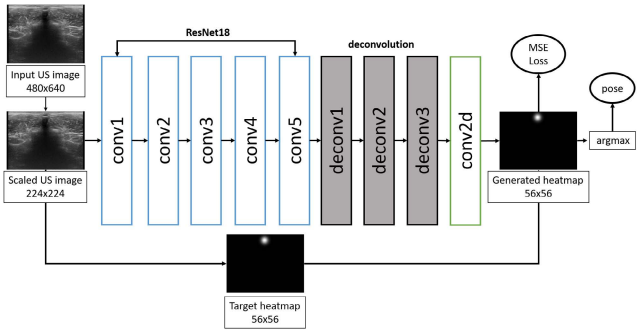


FIGURE 6. Proposed spinous localization network. ResNet based fully connected network with deconvolutional head. Loss is calculated as mean squared error between predicted and target heatmaps. The final pose is calculated from max intensity of the predicted heatmap.

deconvolutional layers with batch normalization [23] and ReLU activation [24], which act as the decoder to generate the image features heatmaps. Each deconvolutional layer has 256 filters with a 4×4 kernel and a stride of 2. The deconvolutional layers are followed by one 1×1 convolutional layer, which transforms the resulting feature matrix to a final heatmap where each pixel intensity represents the probability of being one of the classes; The size of the final heatmap is 56×56. 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 accuracy of detected spinous process locations is calculated in the same way as a PCK metric used in human pose estimation. The detected spinous process is regarded as correct if the distance between the predicted and target spinous processes is less than a certain value, in this case, 50%. In human pose estimation, the head or the torso size is used for scaling the distance error between the predicted and target joint, because the human figures may appear on different scales in the image. Since for spinous process localization there is only one keypoint, there is no need to account for the scale of the structure, thus we normalize the distance to a fixed value. In our method, we use $0.1 \times$ heatmap size, which represents the average size of the spinous process in the image. In this study, we name this new metric as the percentage of correct keypoints of spinous processes (PCKsp@0.5).

The final location is derived from the maximum intensity pixels of the resulting heatmap. To increase the robustness of the dataset, we use combinations of various image data augmentation techniques, randomly chosen for each image: rotation in the range $[-15°, 15°]$, translation in the range $\pm\frac{1}{3}$ of image width and horizontal flipping. To eliminate false-positive findings, the training and validation datasets were subjected to a 5-fold cross-validation experiment split by human subjects, so the images from the same subject could not appear simultaneously at train and test set. Fig. 5 shows US frame, correspondent target heatmap and an example of prediction for that ultrasound frame.

2) SHADOW-BASED METHOD

One of the straightforward methods for highlighting vertebrae is to segment the shadow below the spinous process. This can

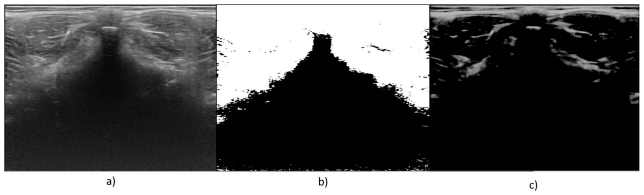


FIGURE 7. Image preprocessing for state-of-the-art methods. a) Original US frame of spinous process b) Binary image for Shadow-based method c) Phase symmetry processed image.

be done by applying conventional image processing methods available within the OpenCV framework. This method imitates the intensity-based methods, such as confidence map [8] or shadow segmentation [6]. Here we do the following steps with the original US frame to detect the spinous process: increase brightness and contrast, apply a threshold to get a binary image (Fig. 7b), filter out small contours, apply dilation to merge contours left into a single big one, find the top point of that contour (which is believed to be a shadow underneath the SP).

3) PHASE SYMMETRY-BASED METHOD

According to the literature review, the state-of-the-art methods for bone enhancement on ultrasound images are the methods based on phase symmetry [2]–[5]. The phase symmetry method helps to highlight the bony features, as Fig. 7c. The phase symmetry calculates image phase information based on the Log-Gabor filter [25]. It looks into the symmetry of intensity values in the image. As result, the phase symmetry image has highlighted edges, the points where the Fourier components are maximally in phase. These edges correlate with maximum image intensities cutting out the specified frequency range.

Parameters for PS were selected experimentally from those listed in the literature and set to: number of wavelet scales 2, number of filter orientations 1, a wavelength of smallest scale filter 25 pix and σ describing log Gabor filter central frequency 0.25 and noise threshold 8. Since it only works based on image intensity, it cannot differentiate properly if it is indeed a bone or a strong reflection from muscular tissue,

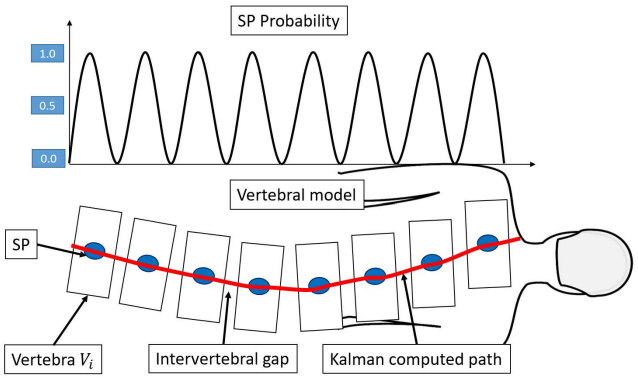


FIGURE 8. Conceptual model of the vertebral and intervertebral gap distribution in the spine and its corresponding expected network output.

thus extra processing is needed to identify the SP position indirectly from other anatomical features, such as lamina. This method applies the following methods from the OpenCV library to PS preprocessed image: apply a threshold to get a binary image, filter out small contours, delete the top part contour which represents the strong reflection from the skin, dilation, find the middle of the detected contours (which represent laminae, sideways features of vertebrae).

F. TRACKING OF SPINAL FEATURES

The sequence of US images for one human subject is called ultrasound sweep. In one sweep, the images of the spinous processes alternate with the intervertebral gaps (see Fig. 8). Each captured ultrasound frame is used as an input to the pre-trained deep learning model (Fig. 6), which outputs the predicted location x_{net} (rescaled to input image size) of the spinous process and its probability (i.e., the confidence of prediction). Since the localization network does not provide meaningful location information for intervertebral gaps (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 of missing location information, we generate a continuous spinal path by using the Kalman filter [26]. The filter estimated locations $\hat{x}_{k-1}$ are fused with the network computed locations x_{net}, whenever they are available (i.e., for the probability of prediction higher than a threshold $thr = 0.5$) resulting in the computed next-frame location $\hat{x}_k$. This method also helps to filter the network output to generate smoother trajectories for the image-based controller of the robot [27].

During an ideal scanning procedure, the path plot will be a straight line positioned at the center of the image coordinates (i.e., half of the image width size). This situation means that the spinous process is kept at the center of the field of view of the ultrasound probe. Any deviations from the center can be considered inaccuracies in the procedure.

G. LEARNING-BASED ROBOTIC NAVIGATION

Ultrasound features are used to compute the velocity command v_x of the robot along the X-axis. The pixel error between the spinous location and the image center is calculated as $e_{img} = \hat{x}_k - w_{img}/2$ (with pixel units), for $w_{img} = 640$px. This error is transformed into meters units (used by the robot controller) as $\delta x = e_{img} \cdot l/w_{img}$, where $l = 0.08$m denotes the probe length. The controller to drive the robot based on the location of the spinous process takes the following standard form: $v_x = -K_x \delta x$, where $K_x > 0$ is a proportional control gain. However, to ensure a smooth motion of the robot, we implement the following first-order filter, which removes any possible outliers produced by network prediction:

$$v_x^k = v_x^{k-1} - \alpha(K_x \delta x^k - v_x^{k-1}) \tag{1}$$

The superscript $*^k$ denotes the discrete-time instance, and $\alpha > 0$ is a control gain. During the scanning task, the robot's vertical motion is commanded with a constant velocity v_y along the Y-axis.

H. FORCE-BASED AUTOMATIC ACOUSTIC COUPLING

The manipulation of US probes over the body involves two main problems: (1) feature-based navigation control (i.e., the methods described above), and (2) stable interaction with deformable tissues; This latter issue is needed to ensure that the probe provides a stable contact with the skin surface to capture US images of acceptable quality. To this end, we implement the following force-based PID controller (which uses the measurements f_z), which drives the probe's velocity v_z along its Z-axis according to the following rule: [28]

$$v_z = -K_p e_f - K_i \int_0^t e_f(\tau)\mathrm{d}\tau - K_d \frac{\mathrm{d}e_f}{\mathrm{d}t}, \tag{2}$$

where $e_f = f_z - f_{ref}$ is the feedback force error, f_{ref} is the reference force to be applied onto the surface. The control gains $K_p, K_i, K_d > 0$ are chosen experimentally according to intrinsic parameters such as the robot's response and extrinsic parameters such as stiffness of the subject's back.

Safety measures are implemented to ensure a safe interaction with human subjects. These force-driven velocity controls are saturated to a safe motion value v_{lim} (as in e.g. [29]) to prevent larger velocities from affecting the stable contact with the skin or causing any discomfort to the subject during the scanning task. Also, a force limiting mode is implemented (at a hardware level) to stop the robot's operation whenever the critical force threshold f_{crit} is exceeded. See Fig. 2.

I. MOMENT-BASED PROBE ORIENTATION CONTROL

We developed a new vision-free strategy to ensure that the ultrasound probe can smoothly follow the curvature of the subject's back on the sagittal plane. This new approach does not rely on any 3D camera measurements to adapt to the curve; It instead uses moment (torque) measurements m_x to automatically adjust the probe into an orientation that minimizes the moment m_x (as distally measured by the force sensor) that is generated by the forces that are lateral to pushing motions. Fig. 9 conceptually depicts the assumed

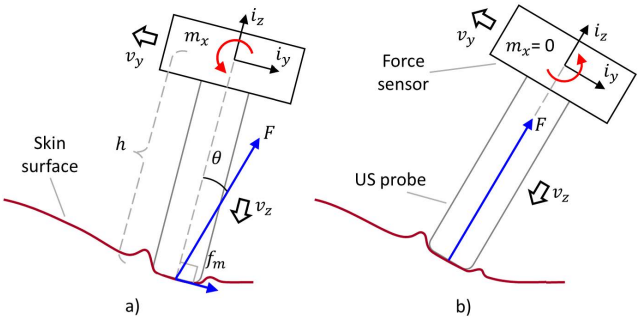


FIGURE 9. Conceptual representation of the proposed moment-based orientation control. (a) Instance when an increasing slope makes the lateral force f_m larger; (b) The probe after the measured moment m_x has been actively minimized by changing its orientation.

FIGURE 10. The probe fixed rotation adjustment based on the shape of the following surface. The probe orientation is maintained normal to the skin surface, the K_{pitch} value is fixed.

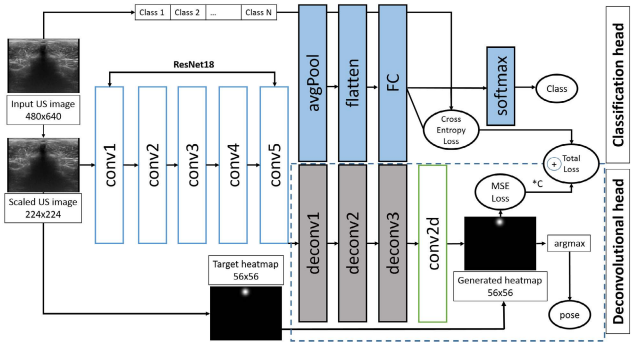


FIGURE 11. Network for Spinal region classification and spinous process localization tasks.

contact model, where h denotes the length of the probe, F the (total) contact force vector applied onto the surface, f_m its lateral force component, and θ the angle between the force vector and the probe's axis. Our proposed model assumes that the probe has already established the desired contact with skin surface $|f_z - f_{ref}| \approx 0$, and moves upwards with a constant speed v_y.

From this model, we can see that the lateral force f_m provides a reliable indication that the surface slope is changing. Such increases/decreases in slope directly result in the generation of a positive/negative moment m_x. This idea suggests that by automatically aligning the probe along the vector F (which implies $m_x = 0$), the robot can *approximately* follow the sagittal curvature of the subject. From the geometric relations in Fig. 9, we can derive $f_m = |F| \sin(\theta)$, which for small values of θ, it can be simplified as:

$$f_m = |F| \cdot \theta \tag{3}$$

We use this equation to compute the moment $m_x = h|F|\theta$, where the applied force along the probe's axis is much larger than the lateral force $f_z \gg |f_m|$, thus, it's reasonable to assume that $|F| \approx f_{ref}$. This enables us to derive the following moment-angle relation $m_x \approx hf_{ref} \cdot \theta$, whose time derivative yields the differential model with angular velocity:

$$\dot{m}_x = (hf_{ref}) \cdot \omega_x \tag{4}$$

The controller to achieve the proposed probe-orienting behaviour is $\omega_x = -K_{pitch} m_x$, for $K_{pitch} > 0$ as a control gain. Substituting ω_x into (4) yields the stable dynamical system $\dot{m}_x = -(hf_{ref} K_{pitch}) m_x$, which asymptotically drives $m_x \to 0$. Note, however, that this approach can only ensure that the ultrasound probe faces the surface at angles close to the surface normal (which corresponds to $\theta \approx 0$). Fig. 10 depicts these moment-based adjustments of the probe's orientation on a human subject.

J. SPINAL REGIONS CLASSIFICATION AND CONTROL

Due to the anatomical spine structure, the lumbar region contrasts with the thoracic and sacrum regions in that it is usually deeper under the layer of fat and muscles, and that has a less curved profile. These differences typically require the sonographer to apply higher pressure and perform fewer rotations with the probe to capture the entire bony structures. Motivated by this human scanning strategy, our method uses a network for spinal regions classification to detect, e.g., when the lumbar region ends and the thoracic one starts. This detection enables to automatically select control parameters, such as the appropriate reference force f_{ref} and the control gain K_{pitch}.

The proposed network has two heads for two different tasks: classification and localization. The first is a deconvolutional head similar to the one depicted in Fig. 6; The second is a classification head used to output the probability of an image belonging to a certain spinal region. The diagram of this multi-task network is depicted in Fig. 11. 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 $C > 0$, which is used to balance the magnitudes of the two losses, $L_{total} = L_{class} + CL_{local}$. The resulting class number where the image frame belongs can be found by applying a softmax layer on the probabilities output from the classification head. The expected output of the network is presented in Fig. 12. The loss for localization task is MSE and for classification is cross-entropy loss.

Fig. 13 shows the proposed application for the spinal regions classification network. The probe in case A is too tilted towards the patient, which might cause the contact to lose and possibly an uncomfortable feeling to the patient. The highly curved lumbar lordosis can cause this issue, which is

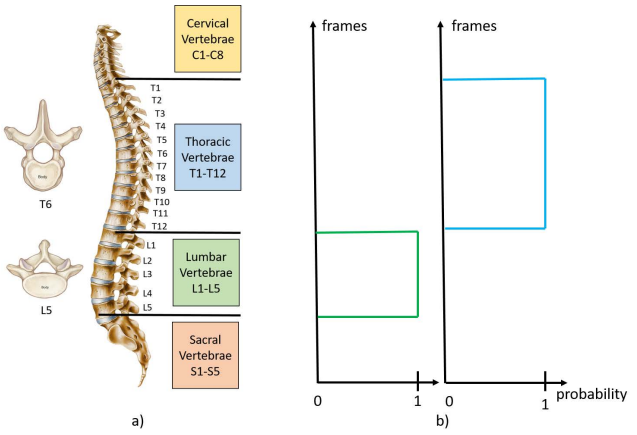


FIGURE 12. a) Spinal regions. b) Expected output of the classification head of the proposed network.

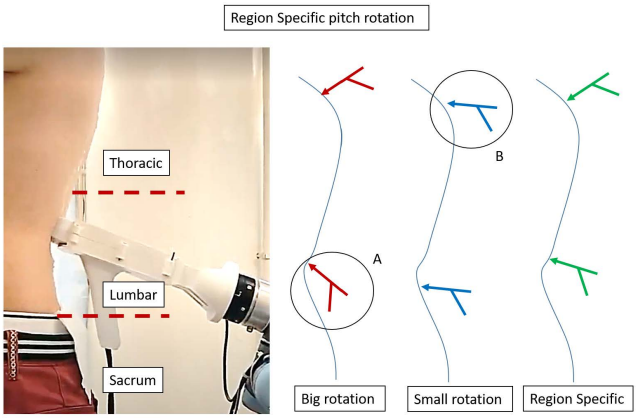


FIGURE 13. Region Specific Rotation. The proposed application of spinal region classification network. The cases A and B indicates problems with fixed K_{pitch} value selection. The region specific scanning proposes the change of K_{pitch} value depending on spinal region.

quite widespread among teenagers. In this case, if the pitch control gain K_{pitch} is low, the probe can freely scan the lumbar region with minimal rotations. However, this potentially can lead to case B for the subjects with the curved upper back. If the value K_{pitch} is low, the probe will not follow the surface correctly and eventually lose contact with the human back. The K_{pitch} values for this gain scheduling-like approach are obtained experimentally on human subjects.

III. RESULTS

A. LOCALIZATION OF SPINAL FEATURES

1) LEARNING-BASED METHOD

The model presented in Fig. 6 was trained for 100 epochs reaching the best validation accuracy at the 10th epoch; The learning rate was 0.0001, the batch size was 12, and Adam [30] optimizer was used. The initial weights for the ResNet [21] backbone were loaded from the publicly available ImageNet pre-trained ResNet18 model. The input images were normalized (to transform the pixels values to floating-point numbers in the range [0, 1]) and resized to 224×224 pixels. The targets were the heatmaps generated from manually labelled images, as shown in Fig. 5a and Fig. 5b. The network was trained on spinous process images collected from 20 subjects and tested on images from 5 subjects. The 5-fold cross-validation was performed by splitting 25 subjects into folds subject-based, thus no images from one subject could appear simultaneously in train and test sets. The resulting accuracy of the 5-fold cross-validation, expressed as PCKsp@0.5, was 0.966 ± 0.027. The mean distance error $e_x = \frac{1}{N} \sum |x_{pred} - x_{target}|$ (where N as the total number of US frames) between predicted and target locations expressed in X image coordinates was 8.0 ± 7.9 pixels, corresponding to 1.0 ± 0.99mm.

2) SHADOW-BASED AND PHASE SYMMETRY-BASED METHODS

The conventional methods were also evaluated on 5 test subjects and resulted in distance errors from predicted locations to ground-truth locations for X image and Y image coordinates are listed in Table 1 with reference to the learning-based method. The values are presented with and without Kalman computed path for X image coordinate. The Kalman computed path reduces the distance error for all methods. The Shadow-based method showed a lower distance error value e_x than the PS method.

B. TRACKING OF SPINAL FEATURES

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 the predicted location. Fig. 14b (left) demonstrates the spinous process presence in the ultrasound sweep, obtained by a manual scan. Red dots correspond to the predicted locations with a confidence of more than 50%; Blue dots denote the corresponding target locations from the manual labeling. For this sample ultrasound sweep, the mean accuracy of spinous process location predictions is 0.995 with a distance mean error of 0.8 mm.

Although predictions are made with high accuracy, there are still some noisy location results, which poses complications for robot control. Thus, the Kalman filter was implemented for real-time prediction of the next frame location of the spinous process; This is done by taking new measurements as input when the prediction accuracy is higher than the threshold of 50%. The filtered path obtained with our method for a US sweep is shown in Fig. 14b (right). The filter's parameters (viz. noise covariance for process $Q = 0.5$ and measurements $R = 500$) were chosen experimentally. Before filtering, the mean accuracy on ultrasound sweeps of 5 test subjects is 0.89 and the mean distance error is 3.3 mm.

The tracking result for the same subject, where the spinous process is detected with the Shadow-based method is presented in Fig. 15. The X mean distance error from predicted

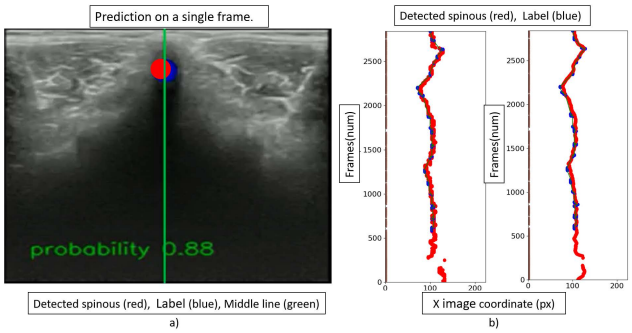


FIGURE 14. FCN based method. The results of the spinous detection on a single frame (a) and the resulted trajectory (b) Left. Real-time detection on a sweep (a sequence of US images obtained from human scan). Blue—labels, red—detected points with threshold of 50%. Right. The resulted Kalman computed path.

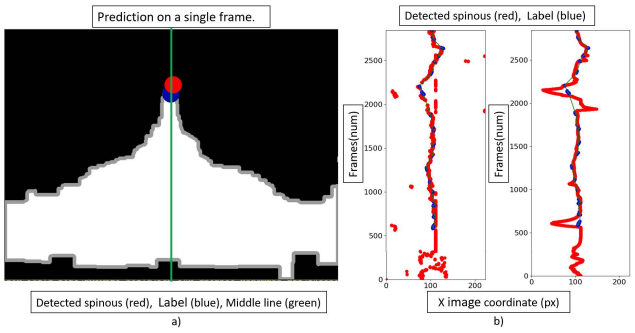


FIGURE 15. Shadow-based method. The results of the spinous detection on a single frame (a) and the resulted trajectory (b) Left. Real-time detection on a sweep (a sequence of US images obtained from human scan). Blue—labels, red—detected points with threshold of 50%. Right. The resulted Kalman computed path.

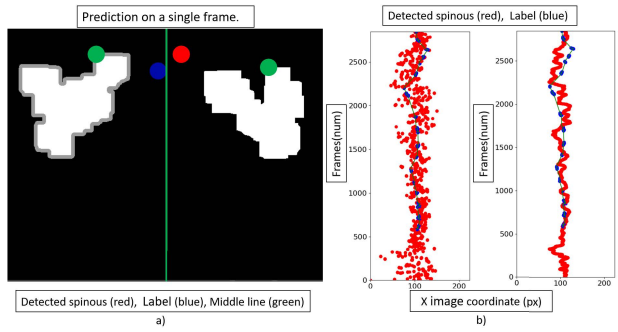


FIGURE 16. Phase symmetry-based method. The results of the spinous detection on a single frame (a) and the resulted trajectory (b) Left. Real-time detection on a sweep (a sequence of US images obtained from human scan). Blue—labels, red—detected points with threshold of 50%. Right. The resulted Kalman computed path.

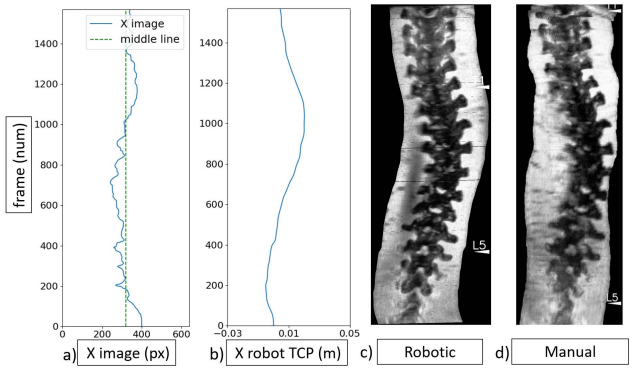


FIGURE 17. Spinous process tracking results on phantom. a) The robot trajectory (m) on the axis with US-images control. b) Spinous process location at US image (pixels) frame by frame during the scan. c) Resulted reconstruction of the phantom's spine from robotic scan and d) from manual scan.

to target spinous process was 4.75 ± 9.5 mm and Kalman processed was 3.4 ± 5.9 mm.

The tracking result for the spinous process detected with the Phase symmetry-based method is presented in Fig. 16. The X mean distance error from predicted to target spinous process was 4.5 ± 3.9 mm and for Kalman filtered 3.6 ± 2.9 mm.

C. EXPERIMENTS WITH PHANTOM AND HUMAN SUBJECTS

The ultrasound system used for manual scanning is a Scolioscan Air platform [31], which consists of a USB ultrasound probe and a tablet that receives images and coordinates from the probe for spinal 3D reconstruction. The proposed control approach was tested on a spinal phantom. Since the US images of the phantom are different from those of a real human spine, a separate model was obtained by training the same network as in Fig. 6 on 1,076 spinous process images obtained with the phantom spine model (this phantom model was only used to safely test the control algorithm). Fig. 17 shows the results of phantom robotic scanning with speed of $v_y = 0.004$ m/s, velocity output clipping of $v_{lim} = 0.05$, reference force $f_{ref} = 7N$, PID control $K_p = 0.0004$ and tilt coefficient $K_{pitch} = 0.02$. Kalman filter parameters were $Q = 0.5, R = 500$. The spinous process location mean is $e_{mean} = 28.6$px (3.5 mm) with $STD = 13.5$px (1.7 mm) for robotic approach.

Fig. 18 shows the results of the robotic scanning for the human subject with mild scoliosis (i.e., with a deformity angle of less than 25 degrees). The plots compare the spinous paths of both robotic and manual scannings. The path for the manual scanning was calculated (with the prediction network) after the manual procedure was performed, while for robotic approach the path was used for real-time navigation. The robotic scanning settings were the following: speed of $v_y = 0.004$m/s, velocity output clipping of $v_{lim} = 0.002$m/s, force setpoint $f_{ref} = 15$N (the force range was chosen according to our previous work on spinal scanning [22] where $f_{ref} = 15$N was selected for subjects with BMI higher than 23), PID control $K_p = 0.0003$, $K_d = K_i = 0.00003$ and tilt coefficient $K_{pitch} = 0.02$. 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 $e_{mean} = \frac{1}{N}\sum(x_{net} - w_{img}/2)$

TABLE 1. Comparison of the methods for spinous process detection, the distance error between predicted and target spinous process across 5 test subjects: FCN — proposed network, PS — phase symmetry-based, and shadow-based.

	FCN (+Kalman)	PS (+Kalman)	Shadow-based (+Kalman)
PCK	0.966 ± 0.027	0.148 ± 0.094	0.095 ± 0.049
Distance error X, e_x, mm	$1.02 \pm 0.99 (0.7 \pm 0.83)$	$6.19 \pm 5.07 (4.19 \pm 3.38)$	$4.10 \pm 9.14 (3.10 \pm 4.51)$
Distance error Y, e_y, mm	1.09 ± 0.99	5.59 ± 3.40	11.36 ± 7.11
Time for one frame, sec	0.02 ± 0.28	0.45 ± 0.034	0.44 ± 0.02

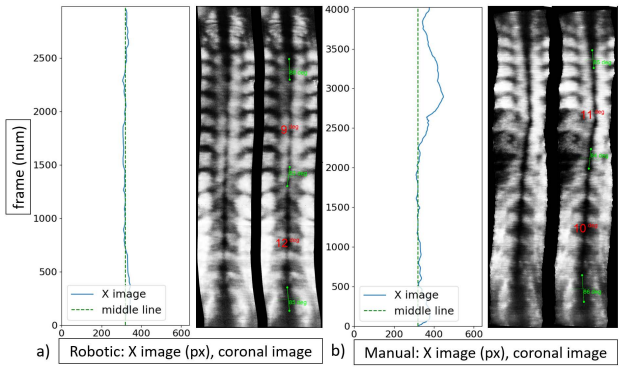


FIGURE 18. Spinous process tracking results on human with mild scoliosis. Kalman computed path (in pixels) frame by frame during the scan, where the green dash line is a middle line of the image. Resulted reconstruction of the human's spine from images collected during a) robotic scanning and b) manual scanning.

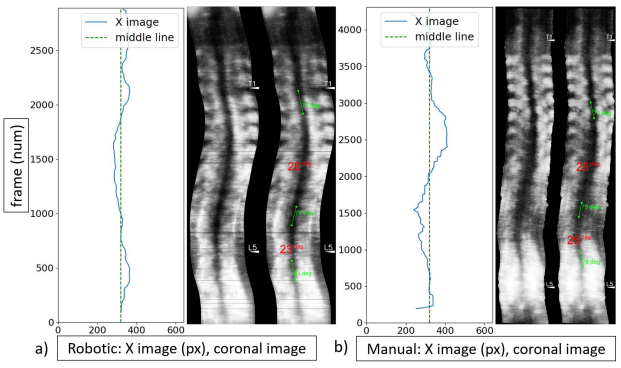


FIGURE 19. Spinous process tracking results on human with moderate Scoliosis. Kalman computed path (in pixels) frame by frame during the scan, where the green dash line is a middle line of the image. Resulted reconstruction of the human's spine from images collected during a) robotic scanning and b) manual scanning.

(for N as the total number of frame in a scan). For this case the $e_{mean} = 7.8$px (1.0 mm) with $STD = 6.5$px (0.8 mm) for robotic approach and $e_{mean} = 36.8$px (4.6 mm) with $STD = 36.9$px (4.6 mm) for manual scanning.

We also evaluated the system with human subjects with moderate scoliosis (i.e., with a deformity angle of more than 25 degrees), see Fig. 19. For this case the $e_{mean} = 22.6$px (2.8 mm) with $STD = 14.7$px (1.8 mm) for robotic approach and $e_{mean} = 34.7$px (4.3 mm) with $STD = 31.5$px (3.9 mm) for manual scanning.

D. ABLATION STUDY

An ablation study is carried out to investigate the contribution of various parts to the overall performance of the scanning robotic ultrasound system. All system modules, including spinous detection, Kalman computed path and moment-based probe orientation control, are included in the overall proposed method. According to offline experiments (section III-A), the spinous detection method based on FCN was found to be more accurate and faster than the other methods, so it will be used in future online experiments. For the ablation study, we consequently turn off one part of the system and perform scanning one by one without spinous detection, without Kalman computed path, and without pitch rotation. During scanning, we compute the distance of the actual spinous process from the image's middle line; the results are shown in Table 2. We've investigated the performance of different methods with excluded modules on a human phantom (Fig. 20) and on human volunteer (Fig. 21).

E. SPINAL REGIONS CLASSIFICATION AND CONTROL

The region detection network was trained to classify ultrasound spine images into three classes: Sacrum, lumbar, and thoracic. Together with the class prediction, the network outputs a 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 a dataset containing US images of the spinous process, and sacrum. Images of only spinous processes were used without intervertebral gap images to imitate the approach in Fig. 6. The images of the spinous processes were split into lumbar and thoracic regions by ultrasound experts.

A total of 20 human subjects (25,774 images), 15 for training, and 5 for testing were used. The best model was achieved on 60 epochs with a learning rate of 0.001, batch size of 12, Adam [30] optimizer was used with a learning rate decay of 0.5 every 20 epochs. The weight of losses C for two heads in $L_{total} = L_{class} + CL_{local}$ was experimentally chosen from [1,500,1000,1500]. The best results were with $C = 1500$, which reflected the approximate magnitude ratios between the classification head loss and the deconvolutional head loss.

The spinous probability results from the localization part of the network output, while the probabilities of the spinal

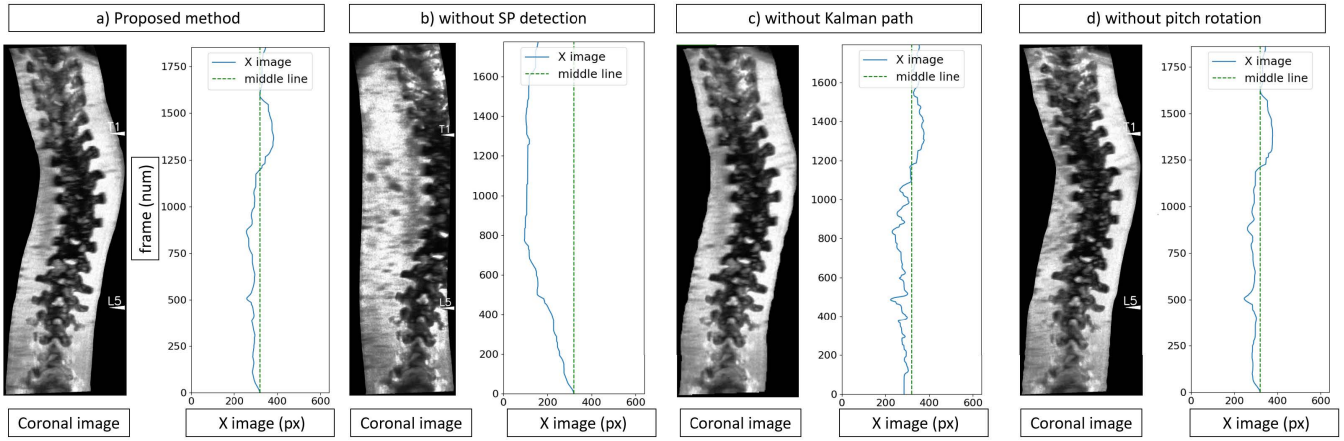


FIGURE 20. Ablation study performed on phantom. Different modes of the system performance: a) proposed method with all parts included, b) without spinous detection with FCN, c) without Kalman computed path, d) without pitch rotation (probe is horizontal to ground).

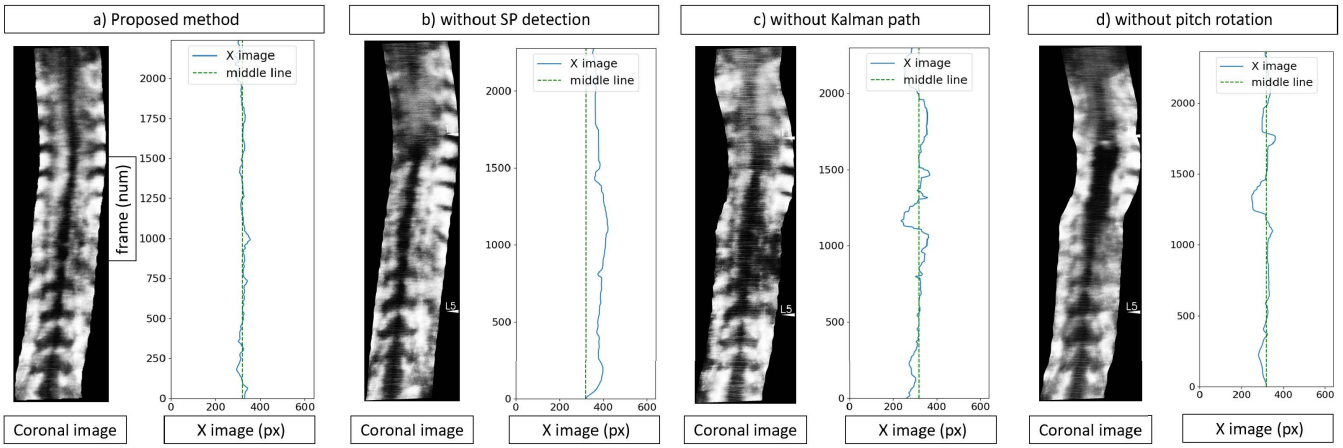


FIGURE 21. Ablation study performed on human subject. Different modes of the system performance: a) proposed method with all parts included, b) without spinous detection with FCN, c) without Kalman computed path, d) without pitch rotation (probe is horizontal to ground).

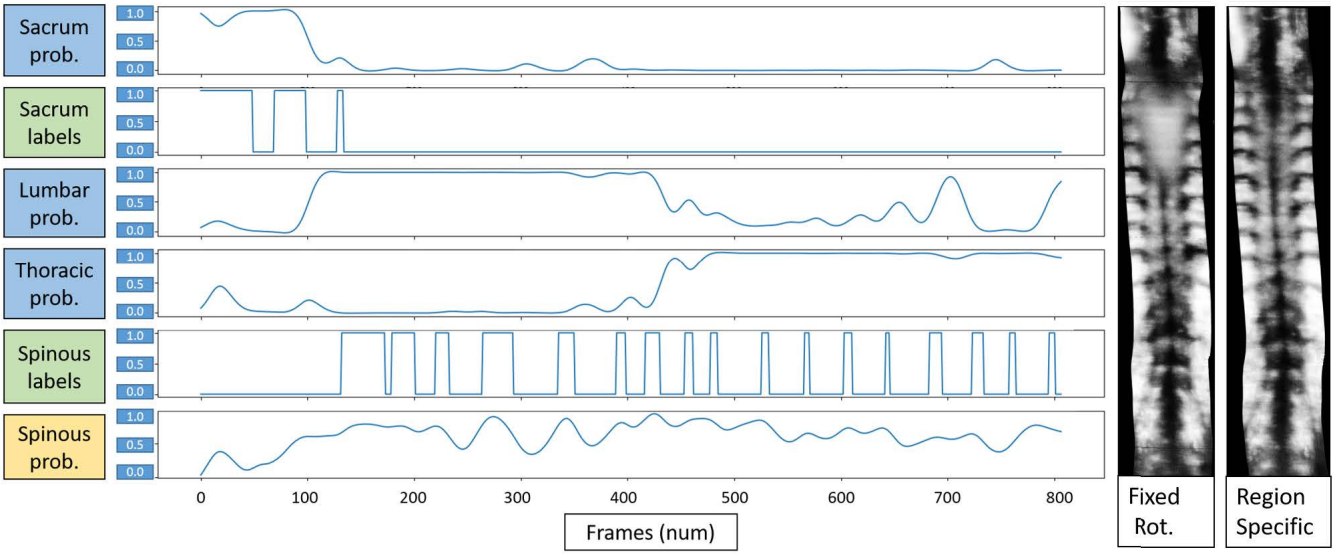


FIGURE 22. Example scan of human subject. Spinal regions network output for 2-head network with classification head of 3 classes. The resulted spinal coronal images with fixed rotation (left) and with region specific rotation (right).

TABLE 2. Ablation study. The scanning system performance is compared for system variations using the distance error e_{mean} (mm) of the spinous process location from the image's central line at each US frame. All modules are included in the proposed method: spinous detection, Kalman computed path, and moment-based probe orientation control. The other methods are variations on the proposed method that do not include one of the modules.

	Spinous detection	Kalman path	Pitch rotation	Distance error, mm (phantom)	Distance error, mm (human)
a	✓	✓	✓	3.88 ± 1.99	1.69 ± 2.13
b	-	✓	✓	20.69 ± 7.93	7.795 ± 2.39
c	✓	-	✓	5.0 ± 3.80	4.59 ± 4.10
d	✓	✓	-	3.91 ± 1.87	2.14 ± 2.09

region are the result of the classification part. These probabilities are presented in a range of 0 to 1, where 1 is a 100% classification confidence. The resulting 3-class models 5-fold cross-validation accuracy on the test set of 5 subjects is 0.94 ± 0.01 for the classification task and 0.97 ± 0.01 for the localization task, see Fig. 22.

The region-specific moment-based probe's orientation control was tested on a human subject, where the rotation gain was changed from $K_{pitch} = 0.02$ for the lumbar region to $K_{pitch} = 0.05$ for the thoracic region. The resulted scan for the subject with a curved thoracic part of the spine was compared to a scan on the same subject performed with fixed rotation $K_{pitch} = 0.02$.

IV. DISCUSSION AND CONCLUSION

The spinous process localization algorithm yields excellent accuracy on the test dataset. The method is robust in defining the spinous process location with a mean distance error of 1.0mm across the test set. The Kalman filter works well on filtering the outliers, wrongly detected as spinous process, and preventing the robot from sudden sideways movements. The filter also helps to form the continuous path, by filling in the locations gaps where the predictions on intervertebral gaps give accuracy less than the threshold.

The designed learning-based spinous process localization method was compared to conventional methods based on phase symmetry and shadow. The learning-based method had the smallest distance error from predicted spinous process to actual location across 5 test subjects for both e_x and e_y. The Shadow-based method had smaller e_x than PS, but a much larger e_y than the other methods. The Shadow-based method is prone to predicting the spinous process in the middle of the image frame, which may be problematic for patients with severe scoliosis. Although the predictions mostly followed the labels, there were a few notable outliers. The PS method predicted locations with fewer acute outliers, but it was less specific in general, predicting an area where the spinous process lies rather than the precise location. Because the distance errors for conventional methods during offline experiments were much greater than those for learning-based methods, they were not used in subsequent online experiments.

The results from the real-time phantom scanning study show that the mean spinous process location distance from the middle of the ultrasound image was around 3.5 mm. By visually assessing the tracking results in Fig. 17, it is clear that the coronal image generated from the robotic scan has the spine centered in the field of view of the probe; It also shows that the spinal features are more prominent, and the edges of the image are smoother. Since the phantom represents an ideal spine model, the navigation performance was further assessed on human subjects.

The conducted human real-time tracking experiments with the final system with all modules included show how well the robotic approach can maintain the spinous process in the middle of US frames for participants with mild and moderate scoliosis (Fig. 18 and Fig. 19). 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 a longer time to center the spinous process in the field of view. Expert sonographers verified (through visual assessment) that: (i) the robotic images for both cases look smoother, and (ii) the spinal features are more distinguishable (compared to the manual approach). This latter result can be explained by the optimal force distribution achieved by the robot along the spine. In general, the distance error for human scanning was smaller than for phantom, which can be explained by a much larger and more generalized dataset for the human model.

To investigate the contribution of different parts of the system, an ablation study was conducted. The proposed method yielded the smallest distance error for both phantom and human, as shown in Table 2. The most significant contribution to system performance is spinous process detection, as turning off this module results in the highest error among other method variations. In terms of system performance, the rotation component was discovered to be less important. It did not affect the image quality of the phantom scan, which can be explained by the phantom's flat surface; however, it did affect the quality of the human scan and caused an air gap (black blur area in Fig. 21d) at the thoracic region of the spine.

The region-specific pitch rotation adjustment turned out to be an essential mechanism for the system. Due to the probe's shape, larger rotations cannot be used if the subject has significant lordosis at the lumbar region; Thus, a smaller rotation gain K_{pitch} should be used. However, if the same subject has a highly curved upper back, smaller rotations will not be sufficient, and the probe will lose contact with the subject's back. Thus, region-specific pitch rotation control can significantly improve the quality of the scanning images.

The spinal region's network performs well according to its accuracy. The robotic experiment with region-specific pitch rotation control clearly shows the improvement in the obtained coronal image, Fig. 22. The robot could smoothly follow the spinal curvature, maintaining tight contact with the subject's skin, and avoiding the air gap in the spine's thoracic (upper) region.

When compared to manual scanning by a human operator, the scoliosis assessment scanning with the robotic system showed improved accuracy for all cases of the spine (as judged by the mean distance error e_{mean}). It was also found that the lumbar features on the robotic scan were more distinguishable, which can be attributed to the advantages of force control. The robotic technique can help to do procedures with minimal operator-patient touch, which is critical in cases like pandemics. It assists in automating the technique at a cost comparable to training an experienced US operator (robotic arm 10,000 USD + software), with a six-month labor cost of approximately 15,000 USD for one operator. Even while humans will not be eliminated from the assessment procedure, they will not be required to undertake repetitive tasks that lead to musculoskeletal disorders, but will instead focus on diagnosis.

There are several limitations of the current work. It would be helpful to have a reference image of the subject's spine, such as an X-ray to compare the scoliosis angles obtained by the robotic system with the ground truth. There were no patients in this study with severe scoliosis, where the spinous process may be absent in transverse ultrasound imaging for some vertebrae due to the severe spinal deformity, which would complicate its localization. Furthermore, patients with a large BMI (e.g., greater than 25) have a deeper-laying spine than subjects with the normal BMI range, affecting vertebral visibility and, as a result, spinous process detection.

Future work for the proposed robotic approach includes a complete system validation and reliability assessment in human trials. It is vital to determine user cases where robotic scanning is more practical or in the contrary only the manual approach can be used, opening the path for future system improvements.

REFERENCES

[1] Y.-P. Zheng, T. T.-Y. Lee, K. K.-L. Lai, B. H.-K. Yip, G.-Q. Zhou, W.-W. Jiang, J. C.-W. Cheung, M.-S. Wong, B. K.-W. Ng, J. C.-Y. Cheng, and T.-P. Lam, "A reliability and validity study for scolioscan: A radiation-free scoliosis assessment system using 3D ultrasound imaging," *Scoliosis Spinal Disorders*, vol. 11, no. 1, pp. 1–15, Dec. 2016.

[2] P. M. S. Shajudeen and R. Righetti, "Spine surface detection from local phase-symmetry enhanced ridges in ultrasound images," *Med. Phys.*, vol. 44, no. 11, pp. 5755–5767, 2017.

[3] I. Hacihaliloglu, R. Abugharbieh, A. J. Hodgson, and R. N. Rohling, "Bone surface localization in ultrasound using image phase-based features," *Ultrasound Med. Biol.*, vol. 35, no. 9, pp. 1475–1487, Sep. 2009.

[4] S. Yu, K. Kiong Tan, B. Leong Sng, S. Li, and A. T. H. Sia, "Feature extraction and classification for ultrasound images of lumbar spine with support vector machine," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 4659–4662.

[5] D. Tran and R. N. Rohling, "Automatic detection of lumbar anatomy in ultrasound images of human subjects," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2248–2256, Sep. 2010.

[6] F. Berton, F. Cheriet, M.-C. Mironand, and C. Laporte, "Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images," *Comput. Biol. Med.*, vol. 72, pp. 201–211, May 2016.

[7] M. Villa, G. Dardenne, M. Nasan, H. Letissier, C. Hamitouche, and E. Stindel, "FCN-based approach for the automatic segmentation of bone surfaces in ultrasound images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 11, pp. 1707–1716, Nov. 2018.

[8] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," *Med. Image Anal.*, vol. 16, no. 6, pp. 1101–1112, Aug. 2012.

[9] J. Hetherington, M. Pesteie, V. A. Lessoway, P. Abolmaesumi, and R. N. Rohling, "Identification and tracking of vertebrae in ultrasound using deep networks with unsupervised feature learning," in *SPIE Medical Imaging*, R. J. Webster and B. Fei, Eds. Bellingham, WA, USA: SPIE, Jul. 2017, Art. no. 101350K.

[10] N. Baka, S. Leenstra, and T. V. Walsum, "Ultrasound aided vertebral level localization for lumbar surgery," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2138–2147, Oct. 2017.

[11] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland, Springer, 2016, pp. 483–499.

[12] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 466–481.

[13] A. Nibali, Z. He, S. Morgan, and L. Prendergast, "Numerical coordinate regression with convolutional neural networks," 2018, *arXiv:1801.07372*.

[14] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1653–1660.

[15] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3686–3693.

[16] Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1173–1182, Feb. 2019.

[17] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, and N. Navab, "Towards MRI-based autonomous robotic U.S. acquisitions: A first feasibility study," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 538–548, Feb. 2017.

[18] O. Zettinig, B. Frisch, S. Virga, M. Esposito, A. Rienmüller, B. Meyer, C. Hennersperger, and Y.-M. R. N. Navab, "3D ultrasound registration-based visual servoing for neurosurgical navigation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, pp. 1607–1619, Feb. 2017.

[19] C. W. J. Cheung, G. Q. Zhou, S. Y. Law, T. M. Mak, K. L. Lai, and Y. P. Zheng, "Ultrasound volume projection imaging for assessment of scoliosis," *IEEE Trans. Med. Imag.*, vol. 34, no. 8, pp. 1760–1768, Aug. 2015.

[20] R. C. Brink, S. P. J. Wijdicks, I. N. Tromp, T. P. C. Schlösser, M. C. Kruyt, F. J. A. Beek, and R. M. Castelein, "A reliability and validity study for different coronal angles using ultrasound imaging in adolescent idiopathic scoliosis," *Spine J.*, vol. 18, no. 6, pp. 979–985, Jun. 2018.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[22] M. Tirindelli, M. Victorova, J. Esteban, S. T. Kim, D. Navarro-Alarcon, Y. P. Zheng, and N. Navab, "Force-ultrasound fusion: Bringing spine robotic-US to the next 'level,'" *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5661–5668, Oct. 2020.

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[25] P. Kovesi, "Image features from phase congruency," *J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.

[26] G. Welch and G. Bishop, "An introduction to the Kalman filter," Dept. Comput. Sci., Univ. North Carolina Chapel Hill, Chapel Hill, NC 27599-3175, Tech. Rep., 1995, pp. 11–15.

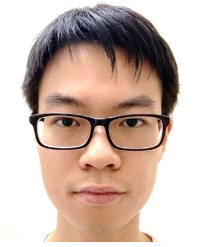
[27] A. Cherubini and D. Navarro-Alarcon, "Sensor-based control for collaborative robots: Fundamentals, challenges, and opportunities," *Frontiers Neurorobot.*, vol. 14, p. 113, Jan. 2021.

[28] M. Victorova, D. Navarro-Alarcon, and Y.-P. Zheng, "3D ultrasound imaging of scoliosis with force-sensitive robotic scanning," in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Feb. 2019, pp. 262–265.

[29] D. Navarro-Alarcon, J. Qi, J. Zhu, and A. Cherubini, "A Lyapunov-stable adaptive method to approximate sensorimotor models for sensor-based control," *Frontiers Neurorobotics*, vol. 14, p. 59, Sep. 2020.
[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[31] K. K.-L. Lai, T. T.-Y. Lee, M. K.-S. Lee, J. C.-H. Hui, and Y.-P. Zheng, "Validation of scolioscan air-portable radiation-free three-dimensional ultrasound imaging assessment system for scoliosis," *Sensors*, vol. 21, no. 8, p. 2858, Apr. 2021.

MARIA VICTOROVA received the B.Sc. degree in medico-technical information technologies from Bauman Moscow State Technical University, Russia, in 2015, the M.Sc. degree in space systems from the Skolkovo Institute of Science and Technology, Skoltech, Russia, in 2017, and the Ph.D. degree in biomedical engineering from The Hong Kong Polytechnic University (PolyU), Hong Kong, in 2022.

During her career she had a visiting appointments at the Technical University of Munich, Germany, and the Higher Technological Institute of Poza Rica, Mexico. She is currently a Research Associate at the Department of Biomedical Engineering, PolyU.



MICHAEL KA-SHING LEE was born in Guangzhou, China, in 1996. He received the B.Sc. degree from The Hong Kong Polytechnic University, Hong Kong, in 2019, where he is currently pursuing the M.Phil. degree in medical imaging. His research interests include ultrasound 3D reconstructions, ultrasound image formation, and prototyping of imaging systems.

DAVID NAVARRO-ALARCON (Senior Member, IEEE) received the M.Sc. degree in robotics from the Centre for Research and Advanced Studies, National Polytechnic Institute of Mexico, in 2009, and the Ph.D. degree in mechanical and automation engineering from The Chinese University of Hong Kong, in 2014.

From 2014 to 2017, he was a Postdoctoral Fellow and then a Research Assistant Professor at the CUHK T Stone Robotics Institute, Hong Kong. Since 2017, he has been with The Hong Kong Polytechnic University, where he is an Assistant Professor at the Department of Mechanical Engineering, and the Principal Investigator of the Robotics and Machine Intelligence Laboratory. He had a visiting appointments at the University of Toulon, France, and the Technical University of Munich, Germany. His current research interests include perceptual robotics and control theory.

YONGPING ZHENG (Senior Member, IEEE) received the B.Sc. and M.Eng. degrees in electronics and information engineering from the University of Science and Technology of China, and the Ph.D. degree in biomedical engineering from The Hong Kong Polytechnic University (PolyU), Hong Kong, in 1997.

After a postdoctoral fellowship at the University of Windsor, Canada, he joined PolyU as an Assistant Professor and was promoted to Professor, in 2008, and the Chair Professor, in 2019. In July 2017, he was appointed as the Henry G. Leong Professor in biomedical engineering. He is currently the Director of the Jockey Club Smart Ageing Hub, and the Research Institute for Smart Ageing, PolyU. His main research interests include biomedical ultrasound instrumentation, soft tissue elasticity measurement and imaging, 3D ultrasound imaging, ultrasound assessment of musculoskeletal tissues, ultrasound image and signal processing, and smart aging technologies.

Prof. Zheng is a fellow of The Hong Kong Institution of Engineers (HK), a Secretary of the World Association of Chinese Biomedical Engineers, from 2017 to 2019, the Past Chair of the Biomedical Engineering Division, HKIE, and an Honorary Advisor of the Hong Kong Medical and Healthcare Device Industry Association (HMHDIA). He serves as the President for the Guangdong Hong Kong Macau Chapter of the International Society of Gerontechnology.