IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 4, 2020, accepted February 15, 2020, date of publication February 28, 2020, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977072

Automatic Markerless Registration and Tracking of the Bone for Computer-Assisted Orthopaedic Surgery

HE LIU AND FERDINANDO RODRIGUEZ Y BAENA, (Member, IEEE)
Mechatronics in Medicine Laboratory, Imperial College London, London SW7 2AZ, U.K.
Corresponding author: Ferdinando Rodriguez y Baena (f.rodriguez@imperial.ac.uk)

This work was supported in part by the China Scholarship Council.

ABSTRACT To achieve a simple and less invasive registration procedure in computer-assisted orthopaedic surgery, we propose an automatic, markerless registration and tracking method based on depth imaging and deep learning. A depth camera is used to continuously capture RGB and depth images of the exposed bone during surgery, and deep neural networks are trained to first localise the surgical target using the RGB image, then segment the target area of the corresponding depth image, from which the surface geometry of the target bone can be extracted. The extracted surface is then compared to a pre-operative model of the same bone for registration. This process can be performed dynamically during the procedure at a rate of 5–6 Hz, without any need for surgeon intervention or invasive optical markers. *Ex vivo* registration experiments were performed on a cadaveric knee, and accuracy measurements against an optically tracked ground truth resulted in a mean translational error of 2.74 mm and a mean rotational error of 6.66°. Our results are the first to describe a promising new way to achieve automatic markerless registration and tracking in computer-assisted orthopaedic surgery, demonstrating that truly seamless registration and tracking of the limb is within reach. Our method reduces invasiveness by removing the need for percutaneous markers. The surgeon is also exempted from inserting markers and collecting registration points manually, which contributes to a more efficient surgical workflow and shorter procedure time in the operating room.

INDEX TERMS Computer-assisted orthopaedic surgery, deep learning, depth imaging, markerless registration.

I. INTRODUCTION

Registration plays an important role in computer-assisted orthopaedic surgery, as it defines the position of the patient with respect to the surgical system so that a pre-operative plan can be correctly aligned with the surgical site. All subsequent steps of the procedure will thus be directly affected by the registration accuracy. Conventionally, two approaches are available to the surgeon. In image-based methods, the surgeon uses a tracked probe to measure the position of a number of points on the target bone, which are compared to their corresponding locations on a plan generated from pre-operative images (e.g. Computed Tomography (CT) or Magnetic Resonance Imaging (MRI)) to calculate the relative spatial transformation. Conversely, in image-free methods, the geometry of the bone surface is scanned using the probe so that a generic model can be morphed onto it for intra-operative planning purposes, avoiding the need for costly pre-operative imaging.

Both registration methods can be defined as 'static', because the registration is only performed once. However, during surgery, the bone will inevitably move, either by the surgeon to adjust the cutting position (in the cm range), or due to cutting or tissue retraction forces (in the mm range). These movements, however small, will cause an error in bone resection if not accounted for. In order to use the registration result throughout the surgery, the target bones have to be rigidly fixed to the surgical system, or markers that can be tracked in real-time have to be inserted into the bones so that the spatial relationship between the system and the bone(s) can be updated continuously without the need for re-registration. The active robotic system ROBODOC (Curexo Technology, Inc.) employs limb fixation, whereas the semi-active orthopaedic robots Mako (Stryker Corp.), Navio (Smith & Nephew PLC) and ROSA Knee (Zimmer

The associate editor coordinating the review of this manuscript and approving it for publication was Anubha Gupta.

Biomet Holdings, Inc.) all use optically tracked markers screwed into the bones.

The tracking methods outlined above all offer high precision, but at the expense of other properties that might affect the outcome of the surgery. Indeed, bone fixation decreases the intra-operative flexibility of the procedure and may even increase the risk of iatrogenic injury and postoperative deep venous thrombosis [1]–[3]. And the invasive procedure of inserting bone fixation pins or marker pins into the bone may cause complications such as infection, vascular and nerve injury, and fracture [3]–[6]. Additionally, in order to complete bone registration, the surgeon needs to manually collect a number of points on the bone surface, which is error-prone, with the accuracy heavily dependant on the surgeon's skill and experience [7]–[9]. Finally, all of the necessary preparations for bone registration (including pin insertion, registration point collection, etc.) will inevitably increase procedure time in the operating room, leading to lower efficiency [2], [10].

If the registration process was fast enough for real-time re-computation, bone fixation or optical markers would no longer be necessary, thus avoiding many of the difficulties described above. By continuously tracking the bone surface, registration can be recomputed on-the-fly, enabling the system to account for any motion of the limb without the need for invasive bone screws. Currently, optical tracking systems used by commercial orthopaedic robots can provide a minimum refresh rate of 20 Hz, which is the lowest requirement for smooth target tracking. Ultrasound can be used intra-operatively to provide images of the bone surface for registration [11], [12], though its speed and convenience for intra-operative use need further investigation. In this context, the development of depth imaging technology, typically adopted by different types of depth cameras, provides new possibilities for bone geometry measurement. Currently, commercial depth cameras can achieve fast ($\geq$ 30 Hz) and accurate ($\leq$ 0.5%) geometry measurement in the form of a high resolution ($\geq$ 640 $\times$ 360) depth image. The captured geometry contains the surface of the bone that can be seen, which can be used to estimate the pose of the target bone.

One of the main challenges with using depth imaging is that it captures all objects in the field of view indiscriminately, whereas only the points that belong to the target bone are useful for registration. Thus, an effective method that can segment the captured depth image is key to the adoption of depth imaging for surface registration in surgery. Image segmentation is an important topic in the field of computer vision, as the partitioning of an image into multiple meaningful segments that can subsequently be extracted and processed is common practice in a variety of application domains. For instance, in medical image analysis, segmentation helps surgeons to separate different cells, tissues, organs, or lesions for diagnosis, treatment planning, etc., and segmentation studies of medical images based on deep learning have achieved satisfactory or even human-level performances [13]–[16], with the potential for a significant reduction in the time, cost, and workload for the clinician.

As a relatively new imaging modality, depth imaging is still rarely applied in surgical scenarios, despite its potential in reconstructing anatomical structures intra-operatively. To the best of our knowledge, the only commercial application of depth imaging to surgical registration is the 7D Surgical System (7D Surgical, Inc.), which captures virtual fiducials of the patient's anatomy using depth cameras so that registration time decreases to less than 20 seconds. However, manual selection of target areas in the depth image is still required, so the registration continues to be 'static', where optical markers are used to keep the registration current. Our previous studies adopted depth imaging for 'dynamic' orthopaedic registration without the need for markers under laboratory conditions [17], [18]. These experiments were performed on bone phantoms, and segmentation of the depth images had to be performed before the registration method could be applied to the real anatomy.

In this paper we present an automatic segmentation method for depth images of a surgical scene, where the segmented depth images are subsequently used for femur registration in computer-assisted knee replacement without the need for invasive markers. Although in knee replacement surgery both femur and tibia need to be registered and resected, tibia registration will undergo a similar process as that for femur registration. Consequently, in this study, we limited our demonstration of markerless registration to the femur, in order to prove the concept and showcase the complete process within a streamlined experimental setup. Our work utilises both RGB and depth images from the depth camera, and deep learning technology to extract useful information from the images for automatic registration. More specifically, our contribution is threefold. First, we create a pixel-wise labelled depth image dataset of the knee anatomy through a normal incision, for femur segmentation training. Second, we apply deep learning to depth image segmentation directly and achieve a mean segmentation accuracy of 87.83% on the testing dataset. Third, we perform online registration using the segmented depth information and demonstrate that depth imaging can be used to obtain markerless and automatic registration in knee surgery.

The paper is organised as follows: in Section II, deep neural networks are developed to localise the target femur and extract the femur surface from the depth images. An automatic labelling method is proposed to label the images from a cadaveric knee for network training. Section III presents the automatic markerless registration method based on the knee segmentation result from Section II, and the registration accuracy is tested in experiments using another cadaveric knee. Experimental results are reported in Section IV, and in Section V, the proposed registration method is discussed in detail, with current limitations summarised and possible solutions identified. A conclusion of the work is given in Section VI and further development is suggested for future work.

II. DEEP NEURAL NETWORK DEVELOPMENT

In this section, deep neural networks are developed to segment depth images of the surgical scene. To reduce the segmentation area, a region of interest (ROI), which is the surgical site of the knee in this study, is first localised by processing the RGB image. Then, the ROI is used to crop the corresponding depth image for segmentation. Both tasks are accomplished using deep learning, and the segmentation output is subsequently used for femur registration.

The reason for using both colour and depth information is twofold. First, depth imaging is still not as developed and stable as RGB imaging, therefore in a depth image, a considerable number of pixels will not have valid values (i.e. their positions cannot be measured), which makes it difficult to globally localise the position of the knee according to the 'broken' depth image. The exponentially increasing measurement noise with respect to distance also diminishes the usability of the whole depth image. Second, RGB imaging is able to provide a robust ROI estimate from the whole scene, but in the smaller scope of the surgical site, the high brightness of the surgical light may compromise the colour features that are useful for segmentation. Bleeding at the surgical site can also complicate the colour conditions, whereas depth imaging is hardly affected. Therefore, the combination of colour and depth information can capitalise on the strengths of both to provide coarse but robust localisation as well as fine segmentation.

A. DATASET CREATION FOR NETWORK TRAINING

A depth image is a map describing the spatial geometry of the scene. Like RGB images, a depth image is also a matrix of pixels, each of which contains three values. Instead of representing colour, the values of each pixel in a depth image are the x, y and z coordinates of that point relative to the depth camera. As depth images and RGB images share the same data structure, the architecture of the artificial neural networks that perform well on RGB images can also be utilised for depth image processing, but because very few studies apply depth imaging to surgical scenarios, there are no labelled datasets of surgical depth images available for segmentation training.

In order to generate a training dataset, we used a commercial depth camera (RealSense D415, Intel Corp.) to scan a cadaveric knee and obtain depth images of its anatomy. RGB images were also collected along with the depth images, and used to train the ROI localisation network. A mechanical rig was designed to hold the cadaveric knee, with a ball joint used to mimic the hip so that the position and angle of the knee could be changed during the experiments. The experimental setup for data collection is shown in Fig. 1.

To facilitate automatic labelling of the femur points in the depth images, the surface points of the femur that could be seen under maximum exposure needed to be collected as a reference. A standard incision was made across the front of the knee to expose the distal femur, then an optical marker M_f that could be tracked by an optical 3D measurement system

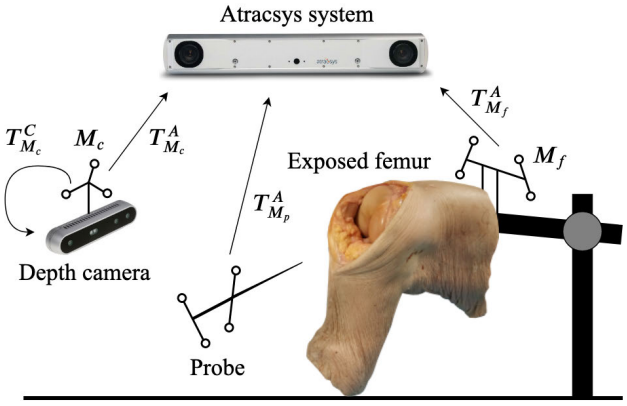


FIGURE 1. Experimental setup for image collection of the knee anatomy.

(fusionTrack 500, Atracsys LLC) was inserted into the femur. A digitising probe was then used to scan the femur surface, and the position of its tip with respect to the femur marker's frame of reference was stored.

After the entire exposed femur surface was scanned using the probe, a point cloud of the femur surface P_f in the femur marker frame was defined, which would be used to label the femur points in the depth images. The depth camera was then used to take depth and RGB images of the cadaveric knee. Another marker, M_c, was attached to the depth camera, and the transformation between the marker frame and the camera frame was computed. The poses of M_c and M_f were measured by the Atracsys system every time a new depth image and a corresponding RGB image were collected. Different positions and angles of the knee were set during the image collection to increase the variability of the data.

More than 2,000 depth images of the knee were collected during the experiment, together with the same number of RGB images, and the poses of M_c and M_f were also recorded for each pair of images. In order to label the points that belong to the femur surface, given the femur points P_f measured by the probe, for each depth image, the femur points can be transformed into the camera frame by

$$P_c = T^C_{M_c} \times (T^A_{M_c})^{-1} \times T^A_{M_f} \times P_f, \tag{1}$$

where $T^C_{M_c}$ is the calibration matrix from the camera marker frame to the camera frame, and $T^A_{M_c}$ and $T^A_{M_f}$ are the poses of the camera marker and the femur marker measured by the Atracsys system, respectively. Theoretically, the points belonging to the femur in the depth image can be labelled by finding the matching points of P_c. However, due to errors existing in the camera-marker calibration, the transformed reference points P_c do not perfectly overlap the femur in the depth image. Therefore, a standard iterative closest point (ICP) [19] algorithm was used to align P_c with the depth image, such that the overlapped points in the depth image could be labelled as the surface points of the femur for segmentation training.

Having labelled the points in the depth images, we augmented the dataset in order to improve training performance.

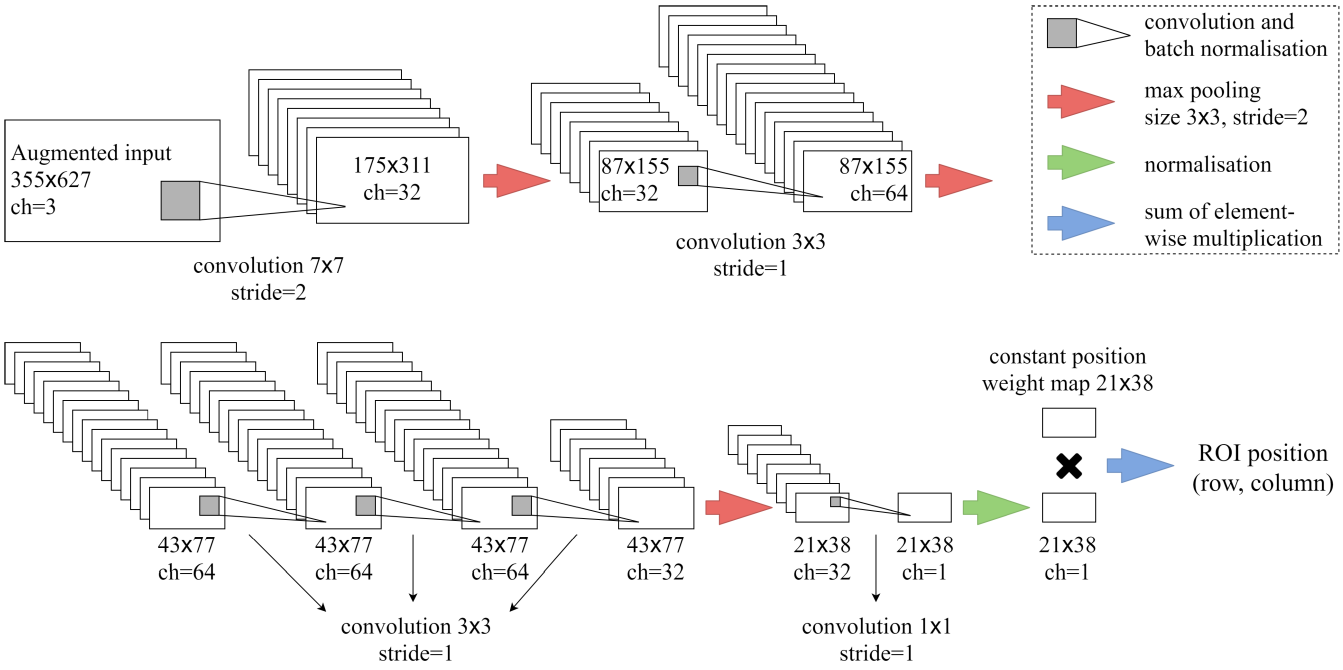


FIGURE 2. Architecture of the ROI localisation network using RGB information.

First, we cropped the depth image to a square shape of size 160 × 160 around the centre of the labelled points, with the cropping centre recorded as the label of the corresponding RGB image for ROI localisation training. Then, the depth images were flipped to increase the size of the dataset. In depth images, pixel position (row, column) and pixel value (x, y, z) are correlated because a pixel in the image is projected from a physical point according to its spatial position. Thus, depth images cannot be augmented by simply flipping the pixel positions, the pixel values also need to be modified. According to the coordinate system of the depth camera, the x-axis points to the right and the y-axis points downwards, so the x values of the pixels change signs if the depth image is flipped horizontally, and y values change signs if the flip is vertical. In this way, a left knee geometry can be produced out of a right one. Rotation is also used to augment the dataset, and for the same reasons as above, the points in the depth image were rotated around the z-axis of the depth camera (pointing forward) by ±90° before the pixels were rotated (anti-)clockwise. For each depth image, the pixel values were multiplied by a random scalar (arbitrarily set between 0.9 and 1.1) in order to represent different knee sizes.

After data augmentation, a dataset of over 10,000 labelled depth images was obtained, which was shuffled and divided into three groups with the ratio of 6:2:2 for network training, validation and testing.

B. NETWORK ARCHITECTURES

Commercial depth cameras normally have a wide field of view that captures a large portion of the environment, so in a typical scenario only a small part of the image belongs to the target bone. In order to decrease the size of the segmentation network and potentially improve its accuracy, we built a localisation network that utilises the RGB information to estimate the ROI position, around which the depth image can be cropped to remove most of the background.

The ROI localisation network has an architecture similar to the AlexNet [20], with five convolutional layers extracting features from the RGB image. However, because we need to preserve the spatial information of the features, instead of using fully connected layers at the end of the network for classification, we apply a 1 × 1 convolutional layer to compress the feature map to one channel and then normalise it. The value of each element in the compressed map represents the probability of the pixel in that position belonging to the ROI. The compressed map is then multiplied element-wise by a pre-defined position weight map to calculate the ROI position. The position weight map has the same size as the final feature map, and each cell in the map has two values representing the relative positions of that cell in the row and the column. The architecture of the ROI localisation network is shown in Fig. 2. Input images are cropped to a given size (355 × 627) to fit the network, then augmented (e.g. flipped vertically and horizontally, with brightness and saturation etc. adjusted randomly) to enlarge the dataset. Batch normalisation [21] is used after each convolutional layer to facilitate network training.

The deep neural network for depth image segmentation adopts the 'U-Net' architecture [15], which is a fully convolutional network with a symmetric 'U' shape, as shown in Fig. 3. The input depth images are randomly cropped to a size of 128 × 128 before being fed to the network, and the output is a 1-channel segmentation map that has the same

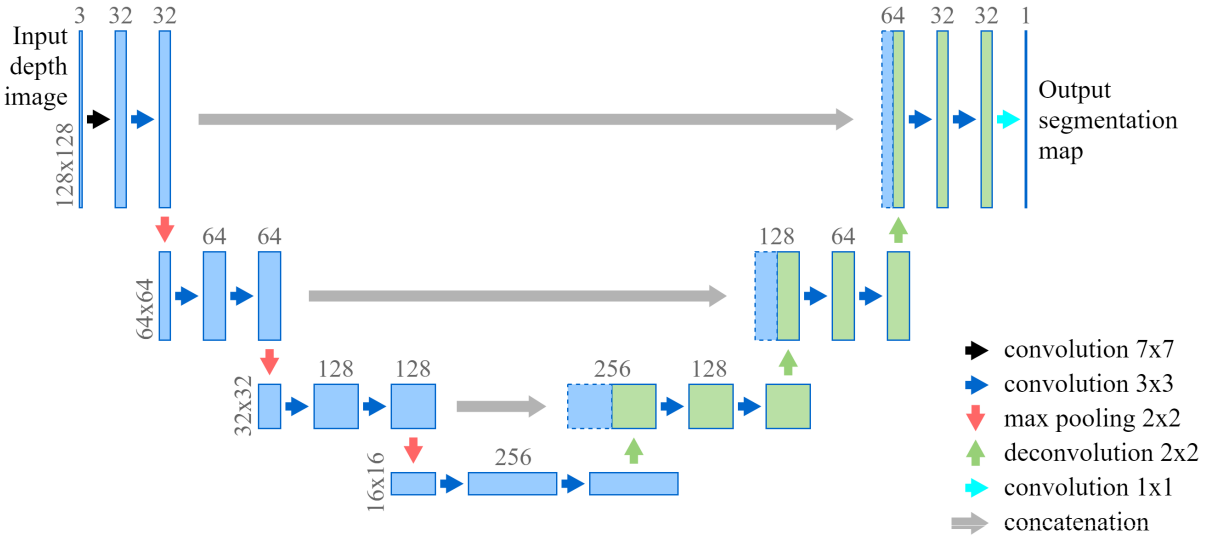


FIGURE 3. Architecture of the depth image segmentation network. The horizontal numbers are the channel numbers of the feature maps, and the vertical numbers are the resolutions of the feature maps.

resolution as the input. On the left side are typical convolutional and pooling layers that increase features and contract resolution, whereas the right side contains deconvolutional layers to increase resolution, which are then concatenated with high resolution features from the left side to assemble a more precise output. The last layer is a 1×1 convolutional layer with a sigmoid activation, mapping all the features of a pixel to a value between 0 and 1, which represents the probability of that pixel belonging to the femur.

C. NETWORK TRAINING AND EVALUATION

Both deep learning networks were implemented using TensorFlow [22], and the Adam optimiser [23] was used for training. For the ROI localisation network, the loss function was defined as the squared distance between the predicted ROI position and the label. To prevent overfitting, dropout and weight regularisation were applied during network training. After training, the testing dataset was used to test the performance of ROI localisation on images that had never been seen by the network. The mean distance between the predicted ROI position and the label was 5.2 (SD: 4.3) pixels, and some of the testing examples of ROI localisation are shown in Fig. 4(a).

The loss function of the depth image segmentation network was defined as the mean of the squared pixel errors:

$$loss = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(p_{ij} - l_{ij}\right)^2 / mn, \tag{2}$$

where m, n are the numbers of rows and columns of the input image, and p_{ij}, l_{ij} are the prediction and the corresponding label of the pixel at row i and column j. Because of the vanishing gradient problem [24] caused by the sigmoid activation in the last layer, it is taxing to train the segmentation network if the parameters are poorly initialised. To facilitate training, first we used the rectified linear unit (ReLU) activation in the last layer and pre-trained the network for a number of epochs to obtain a good initialisation of the parameters, then changed ReLU activation to sigmoid to compute the segmentation map we needed (values between 0 and 1). The segmentation network was trained for 250 epochs before the validation error stopped decreasing.

Different metrics are available to test the image segmentation accuracy, but most of them are designed for binary classification (0 or 1) problems. The values in our generated segmentation map represent the probability of pixels belonging to the bone, which are between 0 and 1 rather than binary. Thus, in order to evaluate segmentation accuracy, we need to either set a threshold to convert the prediction values to binary, or adjust the common metrics for evaluating image segmentation to suit our data format.

Pixel accuracy (PA) is a metric calculating the ratio of pixels that are correctly classified to all pixels:

$$PA = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

where TP, TN, FP and FN represent the pixel counts for true positive (label: 1, prediction: 1), true negative (label: 0, prediction: 0), false positive (label: 0, prediction: 1) and false negative (label: 1, prediction: 0). Given prediction values between 0 and 1, we set a threshold to judge if the prediction is positive (prediction $>$ threshold) or negative (prediction $\leq$ threshold), and derive a weighted pixel accuracy from (3):

$$weighted\_PA = \frac{p(TP) + q(TN)}{p(TP) + q(TN) + p(FP) + q(FN)}, \tag{4}$$

where $p(\cdot)$ means that the contribution of each pixel in that area to the count is the prediction value of that pixel rather than 1, and $q(\cdot)$ means that the contribution is 1 minus the prediction value.

Using pixel accuracy might be misleading when the target area is too small compared to the background because the

TABLE 1. Testing accuracy of depth image segmentation using different metrics.

	Pixel accuracy (%)	IOU (%)
Weighted	99.11 ± 0.63	87.83 ± 9.54
Conventional	98.91 ± 0.69	86.00 ± 9.39

Accuracy is presented in the form of mean $\pm$ standard deviation.

TABLE 2. Confusion matrix of the weighted segmentation results of the testing depth images.

		Actual condition	
		Femur	Non-femur
Predicted condition	Femur	6.99% (TP)	0.42% (FP)
	Non-femur	0.47% (FN)	92.12% (TN)

measure can be biased by a large TN. Therefore, another metric commonly used in image segmentation challenges, called intersection over union (IOU), is used here, which does not account for TN:

$$IOU = \frac{TP}{TP + FP + FN}. \tag{5}$$

Similar to (4), we derive the weighted IOU for our segmentation evaluation:

$$weighted\_IOU = \frac{p(TP)}{p(TP) + p(FP) + q(FN)}, \tag{6}$$

which calculates a weighted pixel count based on the generated pixel probability values.

We used these modified metrics to evaluate the performance of the depth image segmentation, and also converted the prediction values to binary by setting a threshold so that the conventional metrics could be used. The threshold value was set to 0.5 in both cases. The accuracy measured by different metrics is shown in Table 1. Table 2 shows the mean percentages for different fractions in the weighted segmentation results, which highlight a precision of 93.64%, representing the purity, and a recall value of 93.12%, representing the completeness. Some examples of the segmentation results are shown in Fig. 4.

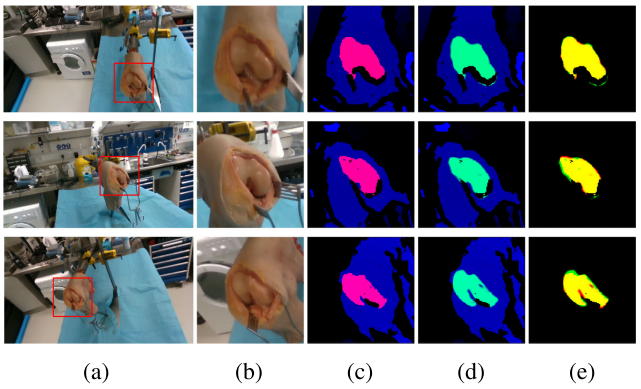


FIGURE 4. Examples of localisation and segmentation results of RGB and depth images. Three rows represent three groups of results from different viewing positions. (a): ROI localisation in RGB images. The centres of the red boxes are the predicted ROI positions, and the size of the red boxes (128 × 128) is used to crop the depth images. (b): RGB images corresponding to the cropped depth images. (c): Depth images (blue) with labelled pixels (ground truths, magenta). (d): Depth images (blue) with predicted femur pixels (cyan). (e): The ground truths (red), the predictions (green) and their overlays (yellow).

III. AUTOMATIC MARKERLESS REGISTRATION

A. MARKERLESS REGISTRATION BASED ON DEPTH IMAGING

After the localisation and segmentation networks were trained with satisfactory accuracy, they were used to process RGB and depth images from the depth camera directly. The localisation network provides the position of the surgical site based on the RGB image, which is then used to crop the corresponding depth image to the required size. The cropped depth image is then fed into the segmentation network to remove surrounding tissues and obtain a clean surface of the target femur, similar to the surface that the surgeon would map out manually using a digitising probe.

Once the distal femur surface has been obtained from the depth image, the pose of the femur can be computed by comparing the acquired surface with a reference model of the femur. Normally this model comes from pre-operative images such as CT or MRI scans that contain the actual surface geometry of the bone. In our case, the reference geometry is acquired by scanning the bone surface using a digitising probe under maximum exposure. The ICP algorithm is an effective and widely used algorithm for precise surface matching in surgical registration [25], [26], but standard ICP requires a large number of iterations to achieve satisfactory convergence. In order to reduce the number of iterations, thus reducing convergence time, a more efficient variant of ICP, the point-to-plane ICP algorithm [27], is adopted. Once initialised with a rough estimate of the limb pose with respect to the camera system, the ICP algorithm will search for corresponding points for each point in the segmented depth image and the reference model, the former representing a subset of the latter in terms of the features available for matching. As per the classical approach, our ICP implementation computes the best pose between these points and uses it to estimate a better candidate pose, which is then applied to the depth image in order to search for better correspondences, until monotonic convergence to a minimum.

Since the segmentation process associates to each point a probability of belonging to the bone, this value is used as a weight of that point when calculating total point-to-plane errors, such that points with higher probability give larger contribution to the ICP pose estimation, which helps further improve the registration accuracy.

In our registration process, the target femur can thus be registered once it is exposed through the surgical incision, and the surgeon does not need to collect registration points manually. Additionally, as the registration can be performed automatically at a fast rate, optical markers are no longer needed, which could contribute to shorter operating times and reduced invasiveness for the patient. The diagram of the markerless

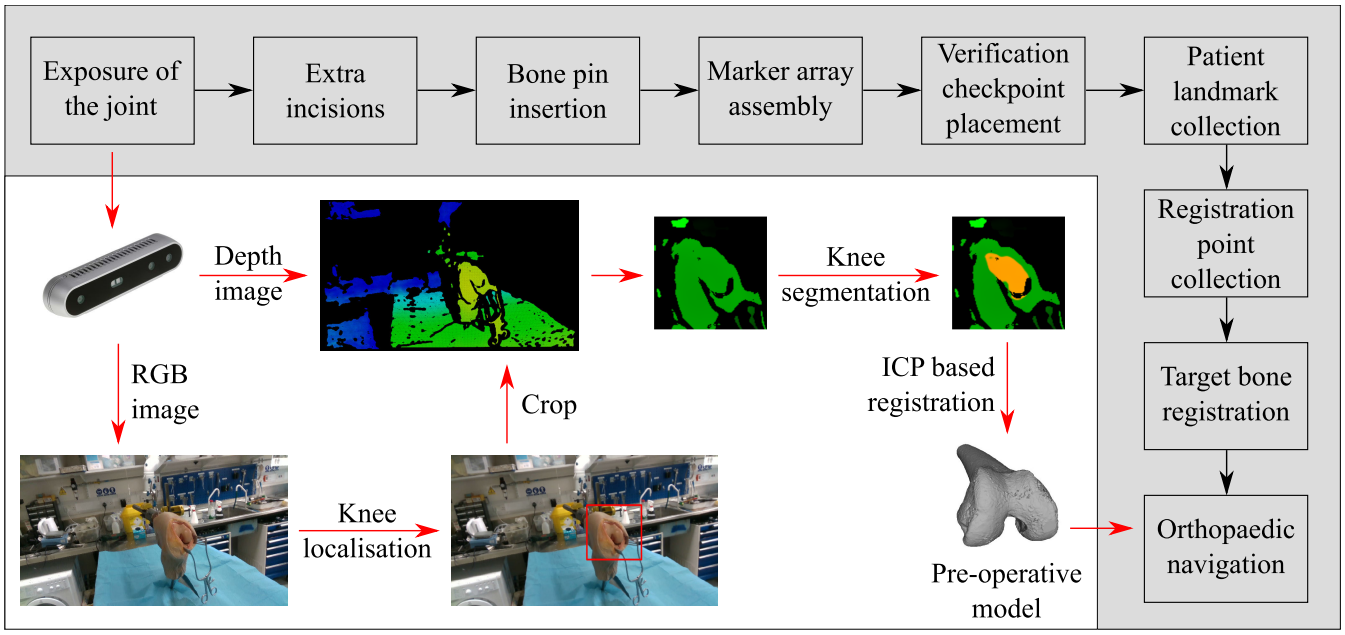


FIGURE 5. Comparison between our automatic markerless registration and conventional orthopaedic registration. The white area highlights the proposed markerless registration workflow, which does not involve surgeon's intervention or invasive markers, whereas the grey area captures the conventional registration process for orthopaedic navigation.

registration process based on depth imaging, alongside a comparison with conventional registration, is shown in Fig. 5.

B. EXPERIMENTS

The main aim of the experiments was to test the registration accuracy of the femur during total knee replacement, without the use of invasive markers. A new cadaveric knee was used in the experiments and a standard incision was made across the front of the knee to expose the distal femur. Although the registration procedure was markerless, in order to measure the accuracy of femur registration, the Atracsys optical tracking system was used again, and the setup was similar to that of the training dataset collection experiment, as shown in Fig. 6. An optical marker was inserted into the femur and the exposed femur surface was scanned using a tracked probe, to obtain a ground truth measurement. The depth camera, with another optical marker attached to it, was then used to capture RGB and depth streams of the exposed knee, from which the femur surface could be localised and extracted.

The experiments included two stages. In stage 1, the depth camera was placed at 40 different positions around the cadaveric knee to measure the pose of the femur while it was fixed in place, and 50 frames were collected at each position to measure the accuracy of static registration. In stage 2, the depth camera was fixed and the knee was moved by hand randomly for 40–50 seconds to evaluate the dynamic tracking ability. The depth camera measured the pose of the femur continuously, and both the depth camera and the target femur were tracked by the Atracsys system for registration accuracy evaluation. The target femur was registered in the depth camera reference frame, then transformed into the Atracsys reference frame using the camera marker tracking results.

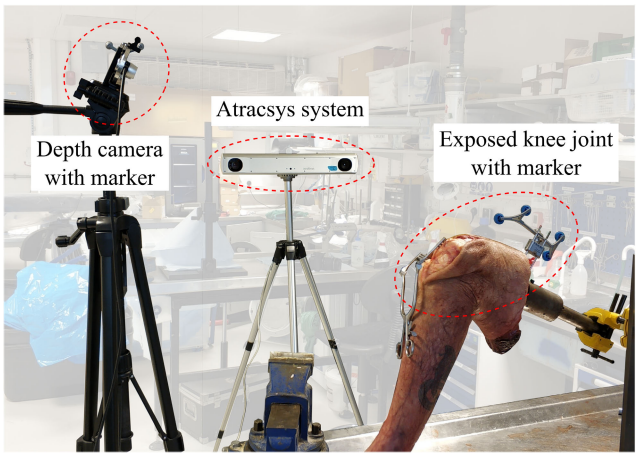


FIGURE 6. Experimental setup for automatic markerless registration on a new cadaveric knee.

To reduce the errors caused by measurement noise, a moving average filter was applied to the transformed femur pose. The ground truth scanned by the probe was also transformed into the Atracsys reference frame to measure the error of markerless registration. Segmentation results of the depth images were saved at random intervals during the experiment, and the time interval between the markerless registration updates was also recorded.

The localisation and segmentation networks were implemented in Python for training, and the architectures and parameters were saved after training so that they could be reloaded and reused in other projects. The software for depth camera control and markerless registration was written in C++, and after the RGB and depth image streams were

TABLE 3. Online segmentation accuracy of the randomly saved depth images.

	Weighted pixel accuracy	Weighted IOU
Accuracy (%)	98.28 ± 0.99	81.87 ± 9.16
Range (%)	96.09–99.23	63.30–92.81

Accuracy is presented in the form of mean $\pm$ standard deviation.

enabled in the depth camera, two parallel threads were created to reload the saved networks and make inferences from the two streams.[1]

The depth camera used in the experiments is the same as the one for training data collection, i.e. the Intel RealSense D415. The whole automatic registration programme was deployed on a computer running Ubuntu 16.04 LTS with an Intel®Core™ i7-4790 processor and 16 gigabytes memory. No external graphic cards were used. Depth image preparation and registration was based on the Point Cloud Library [28], and the Eigen Library [29] was used to facilitate coding of the registration workflow. Result processing and statistical analysis were performed in MATLAB (R2016a, MathWorks, Inc.).

IV. RESULTS

To evaluate the online depth image segmentation accuracy, we randomly saved 20 depth images with the predicted segmentation maps during the experiments. Then, using the same image labelling method as described in Section II-A, we obtained the ground truth segmentation. The segmentation accuracy was evaluated using the weighted pixel accuracy and the weighted IOU, as shown in Table 3.

The registration result provides the pose of the target femur in the depth camera reference frame, which is then transformed into the Atracsys reference frame for comparison with the ground truth, i.e. the pose of the femur scanned by the probe. The overall registration error was measured as the translational and rotational errors measured about the centre of the distal femur, where the rotational error is described as a single screw axis measurement, and in the coronal, sagittal and transverse planes, for easier interpretation. In stage 1, we mainly focused on the static registration accuracy, and the overall errors are shown in Table 4. Box-and-whisker plots for the registration errors in this stage can be seen in Fig. 7.

The aim of stage 2 was to test the dynamic tracking performance of the target femur. We carried out the dynamic tracking of the moving knee three times, and during the motion the knee was always in the field of view of the depth camera. Each motion lasted 40–50 seconds and contained about 250 measurements. The dynamic registration errors of the moving knee are shown in Table 5.

Finally, the time period for the online registration to update was recorded during the experiments. In each period, a series of computations were performed, including preparation of RGB and depth images, identifying ROI from the RGB localisation network, extracting femur surface from the depth image segmentation network, and ICP-based registration. The time recordings when the knee was stationary or moving can be seen in Table 6. From the time recordings, we can see the update rate of the online registration under current hardware and software conditions is about 5–6 Hz.

[1]Thanks to Patrick Wieschollek's package for loading a Python model in C++ on GitHub https://github.com/PatWie/tensorflow-cmake.

TABLE 4. Translational and rotational errors of markerless registration.

	Error	95% CI
3D translational error (mm)	2.74 ± 1.13	[2.69, 2.79]
3D rotational error (°)	6.66 ± 3.07	[6.52, 6.79]
Rotational error in coronal plane (°)[a]	2.09 ± 1.50	[2.02, 2.15]
Rotational error in sagittal plane (°)[b]	-3.50 ± 2.72	[-3.62, -3.38]
Rotational error in transverse plane (°)[c]	-4.33 ± 2.97	[-4.46, -4.20]

Errors are presented in the form of mean $\pm$ standard deviation.
[a] Positive values represent varus and negative values represent valgus.
[b] Positive values represent flexion and negative values represent extension.
[c] Positive values represent retroversion and negative values represent anteversion.

TABLE 5. Dynamic registration accuracy of the knee in motion.

	3D translational error (mm)	3D rotational error (°)
Motion 1	10.87 ± 5.14	9.14 ± 4.46
Motion 2	12.56 ± 6.98	6.89 ± 4.39
Motion 3	12.12 ± 8.28	8.12 ± 7.24

Results are presented in the form of mean $\pm$ standard deviation.

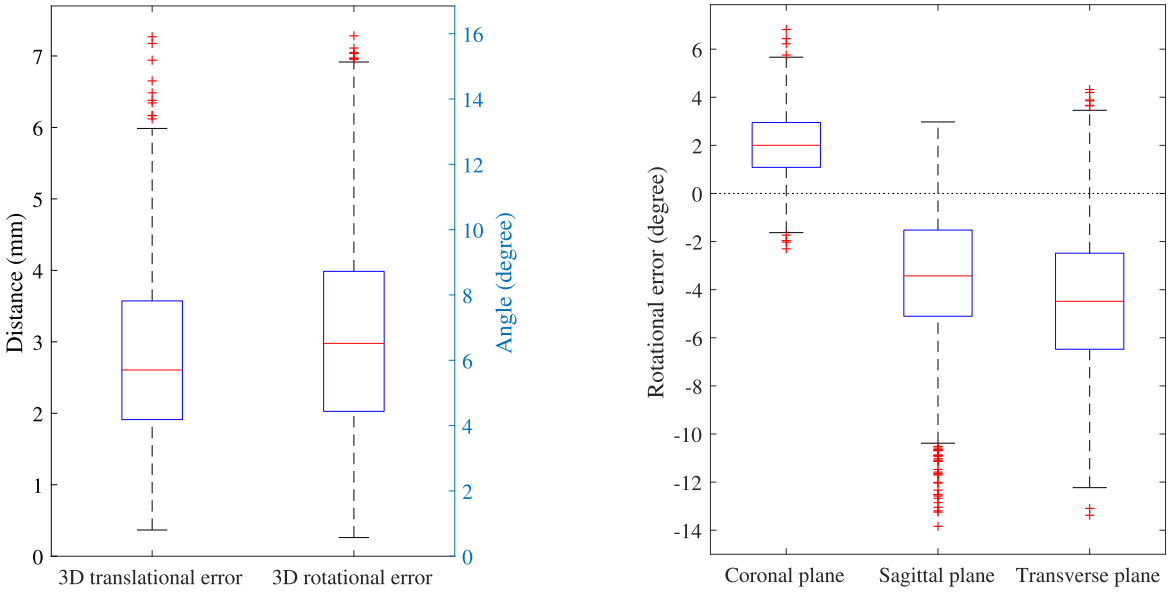
TABLE 6. Computation time for online registration update.

	Fixed knee	Moving knee
Time period (ms)	187.8 ± 14.0	182.7 ± 13.7
95% CI (ms)	[187.2, 188.4]	[181.7, 183.7]

Time period is presented in the form of mean $\pm$ standard deviation.

V. DISCUSSION

This study presents a bone registration and tracking method for orthopaedic surgery that does not require the surgeon's intervention or percutaneous markers. Accuracy of the proposed registration method was measured in experiments, with mean errors of 2.74 mm and 6.66° in translation and rotation, respectively. Although the accuracy, especially rotational accuracy, is currently lower than what can be achieved using conventional intra-operative registration methods based on optical markers [30], [31], the results of this pilot study are still promising, especially considering that the depth camera used in the experiments is not designed for sub-millimetre precision, and its price is much lower than that of the optical tracking system involved in conventional registration methods. Depth cameras are still not as developed as RGB cameras, so noise and obvious outliers are common in the depth images. Outliers exist in small numbers, and most of these can be filtered out by segmentation and thresholding. Measurement noise is more troublesome for the depth camera we use because it is found to distort some areas rather than affecting individual points randomly, which causes bias. This distortion will inevitably lead to registration inaccuracy, which hinders the deployment of our markerless registration



(a) 3D registration errors.

(b) Rotational errors in anatomical planes.

FIGURE 7. Box-and-whisker plots of the registration errors in 3D and in anatomical planes. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered as outliers (approximately $\pm 2.7\sigma$ and 99.3% coverage for a normally distributed dataset). Outliers are plotted individually using the '+' symbol.

method from high-risk applications at this stage. In this study, we focus on providing and validating a new perspective on orthopaedic registration and tracking to inspire more research in this direction, and the registration accuracy can be expected to improve considerably with better depth cameras used in future research.

Compared with the static registration results, the registration errors of the moving knee are larger, especially with translational errors ($p < 0.001$). The main reason could be that the Atracsys system and the depth camera were not strictly synchronised, so the measurements of the 'ground truth' and the registration result did not happen at exactly the same time, possibly causing the large deviation measured in translation. Another possible reason is that the quality of the depth images might decrease when the object is moving, due to technical limitations which are inherent to the depth camera itself. The lower quality of depth images will inevitably introduce larger errors into segmentation and ICP-based registration. Practically, the navigated operations during knee surgery are performed when the leg is approximately still, so the dynamic registration experiments of the moving knee were carried out to illustrate a ''worst case scenario'', which is unlikely to affect the actual operation.

Computational time for the online registration algorithm is also an important aspect of registration performance. With 95% confidence, the registration results can be updated within 190 ms. At the moment, the update rate of the online registration (5–6 Hz) is not satisfactory for clinical use, but we are optimistic about the computation speed because, currently, only the CPU is used to process all the data and make inferences from the trained networks. With more computing power and GPU acceleration, we believe that the computational speed can be drastically increased.

The proposed registration method for knee surgery exploits the power of depth imaging and deep learning, and can benefit both patients and surgeons. For patients, since the bone can be registered and tracked by measuring its surface geometry directly, optical markers would no longer need to be inserted into the bone for real-time tracking, making the surgery less invasive. For surgeons, the surgical procedure could be simplified because no extra operations would be needed to insert the markers. This automatic registration could also exempt them from manually collecting surface points of the target bone, resulting in a significant time saving. The bone could be registered once it is exposed, so a more fluid surgical workflow can be achieved. These advantages will contribute to higher surgical efficiency in the operating room, and a simplified system setup that may eventually lower the cost of surgery.

Apart from being used for registration, the depth image segmentation based on deep learning could also be applied in other clinical tasks. For example, active constraints are useful collaborative control strategies for robot-assisted surgery, which regulates the robotic motion to prevent entering the restricted region [32], [33]. However, the definition of the constraints with respect to the anatomy, especially soft tissues that deform during surgery, is always challenging, as there is no easy way to automatically interpret the patient's anatomy

to define the constraint geometry. Our depth image segmentation provides a possible solution to this problem. Indeed, our method digitises the anatomy in real time and, if properly trained with sufficient data, the neural network could identify different regions of the anatomy dynamically, so that important tissues can be protected. The depth image segmentation method integrates surgeon-like expertise into fast and accurate computer vision, which opens many possibilities for intelligent assistance in surgery.

There are, however, several issues that need to be addressed in future studies before the proposed registration and tracking method can be used clinically. The most straightforward one is to select a medically certified depth camera that is more precise and suitable for the surgical conditions in the operating room, which would alleviate many of the collection inaccuracies identified in this work. Such depth cameras are already available on the market, and have been used in orthopaedic systems such as the 7D Surgical System (7D Surgical, Inc.). In addition, more images of different knee anatomies, ideally in surgical scenarios, will need to be collected to expand the dataset for network training. This will facilitate generalisation of the network inferences, especially on knees with large deformities, and help improve the segmentation accuracy in practical use, thus reducing the registration error.

In our markerless registration, only part of the distal femur surface that is exposed can be measured to register the femur, whereas the proximal end, i.e. the hip centre, is unconstrained, resulting in a large rotational error. In conventional registration methods, however, the position of the hip centre is normally measured, and used to set the mechanical axes of the bones and facilitate registration. It has been demonstrated that, when random noise exists in the surface point measurements, the registration error could be significantly reduced if the hip centre is considered in the ICP algorithm [26]. Therefore, an effective method to estimate the position of the hip centre in real time will need to be explored, using either conventional image processing or deep learning, so that the registration error caused by depth imaging noise can be bounded to a satisfactory range. Alternatively, ultrasound can be used intra-operatively to acquire part of the femur surface that is far away from the knee incision, as in [11], [12], such that the diaphysis direction can be better defined to improve the rotational accuracy.

VI. CONCLUSION

This study proposes a depth image segmentation method based on deep learning, which can be used to achieve automatic markerless registration and tracking of the limb for knee surgery. Deep neural networks were trained using cadaveric data and deployed to perform online segmentation of the depth images from the depth camera, and the accuracy and refresh rate of the proposed registration and tracking method were assessed in *ex vivo* experiments, which demonstrated its effectiveness in both static and dynamic scenarios. Consequently, surgical procedures employing robotics and computer navigation can be simplified considerably, resulting in a more fluent surgical workflow and higher procedure efficiency in the operating room. This method adds intelligence to computer interpretation of anatomical structures without artificial markers, and further extensions can be potentially applied to many other robotic surgical systems.

ACKNOWLEDGMENT

The authors would like to thank Dr. Hadi El Daou for his help with the *ex vivo* trials.

REFERENCES

[1] M. H. L. Liow, P. L. Chin, H. N. Pang, D. K.-J. Tay, and S.-J. Yeo, "Think surgical tsolution-one (Robodoc) total knee arthroplasty," *SICOT-J*, vol. 3, p. 63, Oct. 2017.

[2] A. D. Pearle, P. F. O'Loughlin, and D. O. Kendoff, "Robot-assisted unicompartmental knee arthroplasty," *J. Arthroplasty*, vol. 25, no. 2, pp. 230–237, Feb. 2010.

[3] A. P. Schulz, K. Seide, C. Queitsch, A. von Haugwitz, J. Meiners, B. Kienast, M. Tarabolsi, M. Kammal, and C. Jürgens, "Results of total hip replacement using the robodoc surgical assistant system: Clinical outcome and evaluation of complications for 97 procedures," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 3, no. 4, pp. 301–306, 2008.

[4] S. Gulhane, I. Holloway, and M. Bartlett, "A vascular complication in computer navigated total knee arthroplasty," *Indian J. Orthopaedics*, vol. 47, no. 1, p. 98, 2013.

[5] E. Kamara, Z. P. Berliner, M. S. Hepinstall, and H. J. Cooper, "Pin site complications associated with computer-assisted navigation in hip and knee arthroplasty," *J. Arthroplasty*, vol. 32, no. 9, pp. 2842–2846, Sep. 2017.

[6] R. W. Wysocki, M. B. Sheinkop, W. W. Virkus, and C. J. Della Valle, "Femoral fracture through a previous pin site after computer-assisted total knee arthroplasty," *J. Arthroplasty*, vol. 23, no. 3, pp. 462–465, Apr. 2008.

[7] D. K. Bae and S. J. Song, "Computer assisted navigation in knee arthroplasty," *Clinics Orthopedic Surg.*, vol. 3, no. 4, pp. 259–267, 2011.

[8] R. A. Siston, N. J. Giori, S. B. Goodman, and S. L. Delp, "Surgical navigation for total knee arthroplasty: A perspective," *J. Biomech.*, vol. 40, no. 4, pp. 728–735, Jan. 2007.

[9] G. Zheng, J. Kowal, M. A. González Ballester, M. Caversaccio, and L.-P. Nolte, "(i) Registration techniques for computer navigation," *Current Orthopaedics*, vol. 21, no. 3, pp. 170–179, Jun. 2007.

[10] D. C. Beringer, J. J. Patel, and K. J. Bozic, "An overview of economic issues in computer-assisted total joint arthroplasty," *Clin. Orthopaedics Rel. Res.*, vol. 463, pp. 26–30, Oct. 2007.

[11] P. M. B. Torres, P. J. S. Gonçalves, and J. M. M. Martins, "Robotic motion compensation for bone movement, using ultrasound images," *Ind. Robot, Int. J.*, vol. 42, no. 5, pp. 466–474, Aug. 2015.

[12] P. M. B. Torres, P. J. S. Gonçalves, and J. M. M. Martins, "Robotic system navigation developed for hip resurfacing prosthesis surgery," in *New Trends in Medical and Service Robots. MESROB* (Mechanisms and Machine Science), vol. 48, M. Husty and M. Hofbaur, Eds. Cham, Switzerland: Springer, 2018.

[13] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *J. Digit. Imag.*, vol. 30, no. 4, pp. 449–459, Jun. 2017.

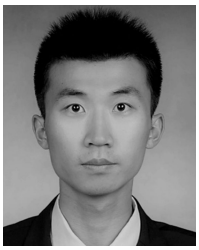
[14] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[15] O. Ronneberger, P. Fischer, and T. Brox "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015.

[16] H. R. Roth, C. Shen, H. Oda, M. Oda, Y. Hayashi, K. Misawa, and K. Mori, "Deep learning and its application to medical image segmentation," *Med. Imag. Technol.*, vol. 36, no. 2, pp. 63–71, Mar. 2018.

[17] H. Liu, E. Auvinet, J. Giles, and F. R. Y. Baena, "Augmented reality based navigation for computer assisted hip resurfacing: A proof of concept study," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1595–1605, May 2018.

[18] H. Liu, S. Bowyer, E. Auvinet, and F. Y. R. Baena, "A smart registration assistant for joint replacement: Concept demonstration," in *Proc. 17th Annu. Meeting Int. Soc. Comput. Assist. Orthopaedic Surg.*, vol. 1, 2017, pp. 189–196.
[19] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 448–456.
[22] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Des. Implement. (OSDI)*, 2016, pp. 265–283.
[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[24] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *Field Guide to Dynamical Recurrent Networks*. Piscataway, NJ, USA: IEEE Press, 2001.
[25] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: A review," *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.
[26] F. R. Y Baena, T. Hawke, and M. Jakopec, "A bounded iterative closest point method for minimally invasive registration of the femur," *Proc. Inst. Mech. Eng. H, J. Eng. Med.*, vol. 227, no. 10, pp. 1135–1144, Aug. 2013.
[27] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vis. Comput.*, vol. 10, no. 3, pp. 145–155, Apr. 1992.
[28] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1–4.
[29] G. Guennebaud and B. Jacob. *Eigen v3*. Accessed: 2010, [Online]. Available: http://eigen.tuxfamily.org
[30] M. Jakopec, F. R. Y. Baena, S. J. Harris, P. Gomes, J. Cobb, and B. L. Davies, "The hands-on orthopaedic robot, 'acrobot': Early clinical trials of total knee replacement surgery," *IEEE Trans. Robot. Autom.*, vol. 19, no. 5, pp. 902–911, Oct. 2003.
[31] S. Nishihara, N. Sugano, M. Ikai, T. Sasama, Y. Tamura, S. Tamura, H. Yoshikawa, and T. Ochi, "Accuracy evaluation of a shape-based registration method for a computer navigation system for total knee arthroplasty," *J. Knee Surg.*, vol. 16, no. 2, pp. 98–105, 2003.
[32] S. Ho, B. Davies, R. Hibberd, and J. Cobb, "Robotic knee surgery-implicit force control strategy with active motion constraint," in *Proc. Euriscon*, 1994, pp. 1235–1248.
[33] B. Davies, M. Jakopec, S. J. Harris, F. Rodriguez y Baena, A. Barrett, A. Evangelidis, P. Gomes, J. Henckel, and J. Cobb, "Active-constraint robotics for surgery," *Proc. IEEE*, vol. 94, no. 9, pp. 1696–1704, Sep. 2006.



HE LIU was born in Jiamusi, China, in 1990. He received the B.Eng. degree in mechanical design manufacturing and automation and the M.Sc. degree in mechatronic engineering from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering with the Mechatronics in Medicine Laboratory, Imperial College London, U.K.

FERDINANDO RODRIGUEZ Y BAENA (Member, IEEE) received the M.Eng. degree in mechatronics and manufacturing systems engineering from King's College London, U.K., in 2000, and the Ph.D. degree in medical robotics from Imperial College London, in 2004.

He is currently a Professor in medical robotics with the Department of Mechanical Engineering, Imperial College London, where he leads the Mechatronics in Medicine Laboratory. His research interests include mechatronic systems for diagnostics, surgical training, and surgical intervention.