2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Robotic Craniomaxillofacial Osteotomy System Using Acoustic 3D Registration*

Jiayu Zhu[1], Runzhe Han[2], Mengning Yuan[3], Bimeng Jie[3], Shanshan Du[1], Yang He[3], Runshi Zhang[1], and Junchen Wang[1]

Abstract— **Osteotomy holds a pivotal position among the fundamental procedures in craniomaxillofacial (CMF) surgery. However, there are inherent challenges and risks associated with ensuring the recuperation of occlusion, safeguarding the facial nerves and blood vessels, as well as preserving facial aesthetics. In this study, a non-invasive image-to-patient registration method for navigation/robotic CMF surgery based on intraoperative freehand ultrasound (US) 3D reconstruction is proposed. Building upon this, a CMF osteotomy robotic system with compliant human-robot interaction and osteotomy trajectory planning was devised. In the freehand US 3D reconstruction and registration experiments, the registration errors for human volunteers and phantoms were consistently less than 1 mm. In robot osteotomy experiments based on the resulting registration, the average osteotomy error was below 1.5 mm. The proposed US 3D reconstruction based registration method is non-invasive and radiation-free, and shows the promising accuracy which is suitable for CMF robotic or navigation systems.**

I. INTRODUCTION

Craniomaxillofacial (CMF) surgery is acknowledged as one of the most intricate surgical fields in medicine. The complexity of the vascular and neurological anatomy of the CMF region, the narrow space, the limited flexibility of instruments, and the obstructed view significantly increase the difficulty and risk of osteotomy. For these reasons, osteotomy requires a high level of clinical experience. Over the last decade, digital technology has developed rapidly in the field of CMF surgery, greatly increasing the accuracy of surgical procedures and reducing post-operative residual deformities.

Orthognathic surgery is a sub-discipline of CMF surgery, in which LeFort I maxillary osteotomy is one of the three basic orthognathic operation types. It is used to move the maxilla in different directions to improve the patient's appearance and correct the occlusal relationship, and accurate osteotomy is the key to the success of the surgery. However, due to the rich blood flow of the maxilla and its close relationship with the surrounding anatomy, osteotomy requires a high level of clinical experience of the surgeon. With the widespread use of digital technology in CMF surgery, precise digital design before surgery has become a routine process, but how to reproduce the preoperative design scheme into surgical operation is a difficult problem in front of the surgeons.

The commonly used clinical approach to executing the preoperative plan involves designing specific surgical templates. However, this introduces additional design steps and can lead to larger incisions and additional trauma. As for robotic surgery, with its high positional accuracy, have the potential to replace the surgical templates and minimize the trauma. Several studies have been conducted to validate the use of robot systems in CMF surgery [1], [2].

Researchers have developed several robot-assisted systems for CMF surgery. Lin et al. developed a surgical robot system that employed a marker, affixed to one side of the patient's mandible with nails, and utilized the AR Toolkit program for registration and tracking [3]. Han et al. utilized an occlusal splint and a six degrees-of-freedom (DoF) mechanical arm to achieve maxillary repositioning in orthognathic surgery [4]. Similarly, Shao et al. leveraged augmented reality (AR) technology and a six DoF robot to assist surgeons in performing osteotomy for mandibular reconstruction surgery [5]. Ma et al. and Deng et al. employed optical cameras to capture six DOF pose of the teeth, which was then registered with CT data [6], [7]. In addition to these approaches, there are other commonly used surgical navigation registration methods. These include directly using bone-implanted titanium screws [8], [9], [10], skin-affixed markers [11], and facial scanning methods employing lasers [12], [13].

While the existing methodologies have yielded relatively satisfactory outcomes, there are four aspects that can be improved. Firstly, the use of nails and screws is invasive, potentially leading to unnecessary trauma. Secondly, employing a splint incurs additional time and costs due to its fabrication. Thirdly, markerless registration techniques that capture the dental structure with a camera risk losing the tracking target amidst surgical bleeding. Lastly, methods utilizing facial markers or laser scanning may suffer diminished accuracy due to soft tissue deformation, such as postoperative swelling.

In this study, we leveraged the nnU-Net neural network [14] to segment temporal 2D CMF US images. Subsequently, a combination of Pixel Nearest Neighbor (PNN) algorithm [15] and Bezier-based voxel filling technique [16] was utilized for freehand 3D US reconstruction. Preoperative

*This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFC2405401 and in part by the Natural Science Foundation of China under Grants 62173014 and U22A2051 and in part by the Natural Science Foundation of Beijing Municipality under Grant L232037. (Corresponding author: Junchen Wang)

[1]Jiayu Zhu, Shanshan Du, Runshi Zhang, and Junchen Wang are with the School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China (J. Wang: wangjunchen@buaa.edu.cn)

[2]Runzhe Han is with the School of General Engineering, Beihang University, Beijing 100191, China

[3]Mengning Yuan, Bimeng Jie and Yang He are with the Department of Oral and Maxillofacial Surgery, Peking University School and Hospital of Stomatology, Beijing, 100081, China

planning of the osteotomy lines was performed on the patients' cone beam computed tomography (CBCT) scans, which were then registered with the intraoperative 3D US reconstruction. This was followed by translating the osteotomy plan to the robot system. By incorporating an admittance control method, surgeons could easily position the robot near the intended osteotomy site. Following this, the robot precisely executed the osteotomy in accordance with the preoperative plan. To validate the system's reliability and effectiveness, comprehensive experiments were conducted on both phantoms and human volunteers.

II. System Overview and Methods

A. System Overview

Fig. 1 illustrates the schematic diagram of the acoustic registration-based robotic system designed for CMF osteotomy surgery. This system integrates intraoperative freehand 3D US reconstruction with the patient's preoperative CBCT for surgical navigation. The CMF bones are segmented from the CBCT data volume employing a deep learning method approach as detailed in our recently published article [17]. To capture 2D US images of the patient's maxilla, a portable US device is utilized. The 2D US image coordinate frame $\{\mathrm{I}\}$ is associated with the US probe. The frame $\{\mathrm{U}\}$ originates from the optical marker attached to the US probe, with the relationship $\boldsymbol{T}_{\mathrm{UI}}$ between the image $\{\mathrm{I}\}$ and the marker $\{\mathrm{U}\}$ established through the probe calibration [18]. An additional optical marker represented by the reference frame $\{\mathrm{H}\}$ is securely fastened to the patient's head using a bandage, allowing for optical tracking of the patient with respect to the optical device represented by the frame $\{\mathrm{W}\}$. This setup enables the relative tracking of $\{\mathrm{U}\}$ and $\{\mathrm{H}\}$, minimizing the adverse effect of patient's shift. The 3D US image reconstruction is defined within the coordinate system $\{\mathrm{V}\}$ which is fixed and known with respect to the reference $\{\mathrm{H}\}$. The surgical preoperative planning is performed on the segmented CBCT model with the coordinate system $\{\mathrm{C}\}$. The transformation $\boldsymbol{T}_{\mathrm{VC}}$ between the intraoperative 3D US reconstruction and preoperative CBCT of the patient is achieved by the point cloud registration [19]. Furthermore, to link the task space of the robotic manipulator with the optical tracker, four coordinate systems are defined which are respectively the robot base $\{\mathrm{B}\}$, the end flange $\{\mathrm{F}\}$, the optical marker $\{\mathrm{M}\}$ attached to the flange, and the end-effector $\{\mathrm{T}\}$. The transformation $\boldsymbol{T}_{\mathrm{FM}}$ is derived from the hand-eye calibration, and $\boldsymbol{T}_{\mathrm{MT}}$ is obtained through the calibration of the end-effector, facilitated by the optical tracker.

B. Image Segmentation

As depicted in Fig.1, the 2D US images of the human zygomatic and frontal bones, acquired via US scanning, are segmented and subsequently reconstructed into 3D models. The 3D US reconstruction is then registered with the preoperative CBCT to replicate the preoperative planning accurately. The fidelity of both image segmentation and reconstruction critically influences the precision of subsequent registration and navigation processes. The acceptable registration accuracy is generally considered to be less than 1 mm [20].

US imaging is often hampered by challenges such as noise, artifacts, and low contrast, with the imaging quality further contingent on the operator's proficiency and technique. Consequently, the accurate segmentation of US images presents a significant challenge [21]. To address this, we have employed deep learning approaches for image segmentation. nnU-Net is a variant of U-Net [22], which is widely used in the field of medical image segmentation. nnU-Net has the ability to evaluate the training set and automatically generate network structures, along with excellent image preprocessing algorithms. Leveraging this, we have improved the network's training strategy, focusing on enhancements in two key aspects: the loss function and the learning rate strategy.

The original nnU-Net uses Dice+Cross Entropy loss function, where the Dice loss can be expressed as:

$$loss_{\mathrm{Dice}} = 1 - \frac{2TP}{FP + 2TP + FN} \tag{1}$$

We replaced the Dice loss with the Tversky loss [23]:

$$loss_{\mathrm{Tversky}} = 1 - \frac{TP}{TP + pFP + qFN} \tag{2}$$

Here, TP denotes the number of true positive pixels, FP represents the number of false positive pixels, and FN signifies the number of false negative pixels. By setting $p+q=1$ and adjusting the values of p and q, the proportion of FP and FN in the loss function can be modified. This adjustment allows to selectively bias the training towards either precision or recall, addressing the sample imbalance problem caused by the disproportionately small foreground area in US images.

To expedite training process, we introduced a warm restart strategy [24] to the poly learning rate approach originally utilized in nnU-Net. For further details on implementing US image segmentation with nnU-Net, refer to section III-A.

C. 3D Reconstruction

To acquire US data, surgeons perform freehand US scans, drawing upon their experience. An optical tracker is utilized to collect pose data for each 2D US image, ensuring their proper alignment. Given the intricate contours of the face, scanning certain bone tissues thoroughly can be challenging. In order to achieve the completeness and smoothness of the 3D US reconstruction model for bone surface, we introduce a 3D reconstruction technique that combines image downsampling with Bezier functions, termed as DSBezier.

The algorithm's computation workflow is illustrated in Fig. 2. To reduce reconstruction time, the input 2D images are first downsampled by a factor of 4. Then, the first set of 4 images is selected to construct Bezier curves and to compute the voxel values along the curve's trajectory. Following this computation, the first 2 images in the queue are discarded. Concurrently, new data are enqueued. The process then proceeds with the next set of 4 images, which included 2 images that overlap with the previously used set,

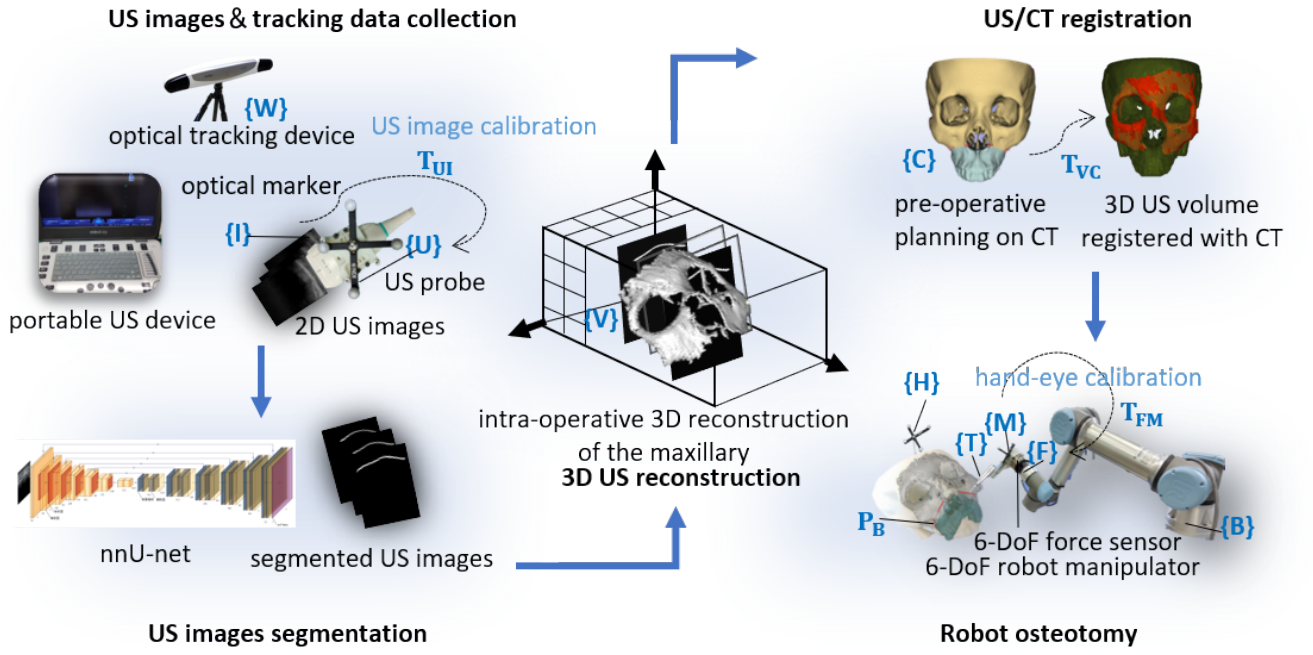

Fig. 1. Schematic diagram of the acoustic registration-based robot system for CMF osteotomy surgery. The system comprises an optical tracker, a portable US device, a US probe equipped with an optical marker {U}, and a 6-DoF robotic manipulator. Attached to the robotic manipulator's flange is a 6-DoF force sensor, an end-effector and an additional optical marker {M}.

for generating new Bezier curves. This cycle continues until the final set of 4 images is processed.

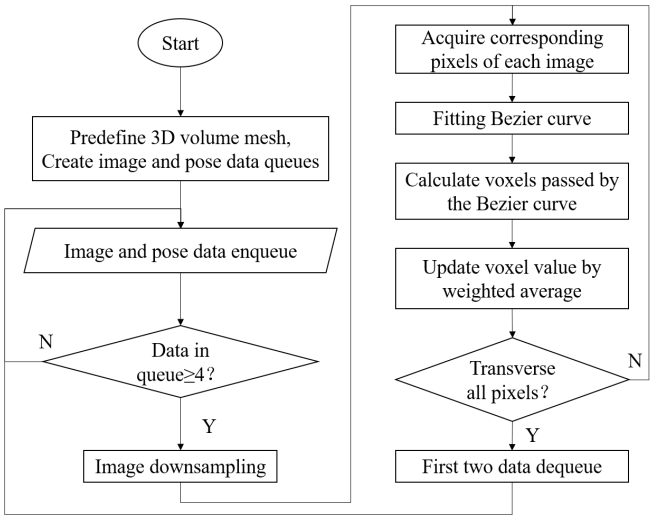

Fig. 2. The workflow of freehand US 3D reconstruction

The coordinate of a pixel expressed in {H} can be calculated by $\boldsymbol{P} = \boldsymbol{T}_{\mathrm{WH}}^{-1}\boldsymbol{T}_{\mathrm{WU}}\boldsymbol{T}_{\mathrm{UI}}[s_x u, s_y v, 0, 1]^{\mathrm{T}}$, where u and v denote the pixel's location in the US image, and s_x and s_y represent the pixel's size. The pixel coordinates and values for 4 images, denoted as $\boldsymbol{P}_i = (x_i, y_i, z_i, v_i)^{\mathrm{T}}, i = 1, 2, 3, 4$ are subsequently utilized to construct a Bezier curve:

$$\boldsymbol{P}(t) = \boldsymbol{P}_1(1-t)^3 + 3\boldsymbol{P}_2 t(1-t)^2 + 3\boldsymbol{P}_3 t^2(1-t) + \boldsymbol{P}_4 t^3 \quad t \in [0,1] \tag{3}$$

By adjusting the parameter t, voxels traversed by the curve can be assigned values employing the PNN method.

D. Robot Intraoperative Control

1) Robotic Osteotomy Trajectory Plan: Upon completion of the intraoperative 3D US reconstruction of the segmented frontal and zygomatic bones, a mesh model is generated using the marching cubes algorithm [25]. The reconstructed bone surfaces is further registered to the patient's preoperative CBCT skull model using the iterative closest point (ICP) method, obtaining the registration matrix $\boldsymbol{T}_{\mathrm{VC}}$. With $\boldsymbol{T}_{\mathrm{HV}}$ specified during reconstruction and $\boldsymbol{T}_{\mathrm{WH}}$ acquired from the optical tracker, the osteotomy plan in {C} can be converted to the intraoperative patient's reference frame {H}.

The osteotomy plan for the end-effector, as defined in the preoperative CBCT, is denoted by $\boldsymbol{T}_{\mathrm{C}}$, comprising orientation $\boldsymbol{R}_{\mathrm{C}}$ and position $\boldsymbol{P}_{\mathrm{C}}$. Through optical tracking, the osteotomy plan expressed in the robot's base $\boldsymbol{T}_{\mathrm{B}}$ can be computed as follows:

$$\boldsymbol{T}_{\mathrm{B}} = \boldsymbol{T}_{\mathrm{BF}}\boldsymbol{T}_{\mathrm{FM}}{\boldsymbol{T}_{\mathrm{WM}}}^{-1}\boldsymbol{T}_{\mathrm{WH}}\boldsymbol{T}_{\mathrm{HV}}\boldsymbol{T}_{\mathrm{VC}}\boldsymbol{T}_{\mathrm{C}} \tag{4}$$

The end-effector's position and orientation are decoupled by:

$$\boldsymbol{R}_{\mathrm{C}}(s) = \boldsymbol{R}_{\mathrm{C,start}}\exp(\log(\boldsymbol{R}_{\mathrm{C,start}}^{\mathrm{T}}\boldsymbol{R}_{\mathrm{C,end}})s) \tag{5}$$

$$\boldsymbol{P}_{\mathrm{C}}(s) = \boldsymbol{P}_{\mathrm{C,start}} + s(\boldsymbol{P}_{\mathrm{C,end}} - \boldsymbol{P}_{\mathrm{C,start}}) \tag{6}$$

where $s(t) \in [0,1]$ is a time scaling function which usually has a form of trapezoidal velocity curve.

2) Admittance Controller for Compliant Human-Robot Interaction: To mitigate the risk of potential collisions in complex surgical settings, admittance control has been implemented. This approach significantly enhances the maneuverability of the surgical robot manipulator, allowing the surgeons to manually guide the robot to the vicinity of the osteotomy site before activating the automatic osteotomy mode.

In our model, the six DoF robotic manipulator is conceptualized as a six-dimensional mass-damping-spring system. We employ screw theory, utilizing twists and wrenches to develop the robot's admittance controller:

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{B}\boldsymbol{\nu} + \boldsymbol{K}\boldsymbol{S}\theta = \boldsymbol{\mathcal{F}}_{\text{ext}} \tag{7}$$

Here, $\boldsymbol{\mathcal{F}}_{\text{ext}}$ denotes the interaction force measured by the six-dimensional force sensor attached to the robot's wrist, $\boldsymbol{\nu} \in \mathbb{R}^6$ represents the robot's velocity, and $\boldsymbol{S}\theta$ signifies the exponential coordinates of the robot's end-effector. The matrices $\boldsymbol{M}$,$\boldsymbol{B}$, and $\boldsymbol{K} \in \mathbb{R}^{6\times6}$ are positive-definite, depicting the virtual mass, damping, and stiffness characteristics that the robot simulates. To preserve compliance with the surgeon's control and eliminate spring-like behavior, a modified admittance controller model is employed:

$$\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{B}\boldsymbol{\nu} = \boldsymbol{\mathcal{F}}_{\text{ext}} \tag{8}$$

By addressing the inverse velocity kinematics, the robot's joint velocity can be calculated, which are commanded to the robot's velocity controller. Let $\boldsymbol{J}_{\text{T}}^{\dagger}$ denote the pseudo inverse of the Jacobian matrix expressed in the end-effector frame $\{\text{T}\}$, the calculation is as follows.

$$\begin{aligned} \dot{\boldsymbol{\nu}} &= \boldsymbol{M}^{-1}(\boldsymbol{\mathcal{F}}_{\text{ext}} - \boldsymbol{B}\boldsymbol{\nu}) \\ \boldsymbol{\nu} &\leftarrow \boldsymbol{\nu} + \dot{\boldsymbol{\nu}} \triangle t \\ \dot{\boldsymbol{\theta}} &= \boldsymbol{J}_{\text{T}}^{\dagger}\boldsymbol{\nu} \end{aligned} \tag{9}$$

III. Experiments and Results

The objective of our experiments was to validate the precision of the robotic system for maxillofacial osteotomy, leveraging intraoperative 3D US reconstruction registered with preoperative CBCT plans. The experiments included: 1) segmentation of 2D US images; 2) assessment of 3D US reconstruction and registration accuracy on both phantoms and human volunteers; and 3) execution of robotic maxillary osteotomy on the phantoms. The equipments used in the experiments included: a portable US imaging system with a linear probe (Mindray M9, Shenzhen, China), a force-torque force sensor (ATI Gamma, Apex, USA), an optical tracker (Polaris, NDI, Canada), and a 6-DoF robot manipulator (UR5e, Universal Robots, Denmark). The algorithms for the US 3D reconstruction, image registration and robot control were implemented in C++ as the software platform developed by our laboratory (https://mrs.buaa.edu.cn/?page_id=342).

A. Image Segmentation

1) Dataset and Setup for Training: A dataset comprising 882 zygomatic US images from 8 patients, annotated by medical professionals from the Peking University School and Hospital of Stomatology, was used to train the nnU-Net. Out of these, 706 images constituted the training set, while 177 images formed the test set. The training process spanned 1000 epochs.

2) Experiments: The enhanced nnU-Net incorporates additional parameters, including p and q in the Tversky loss, alongside the timing of restart and the drop of initial learning rate in the warm restart strategy. After multiple rounds of training, we found the optimal parameters: $p = 0.3$, $q = 0.7$, with a warm restart activated after epoch 250, and a 65% drop of the initial learning rate. Results in Table I show that the improved nnU-Net achieved a Dice score of 0.8405 and outperformed the original nnU-Net across four evaluation metrics: Dice, mIoU, Recall, and Precision.

TABLE I
DICE, MIOU, RECALL AND PRECISION FOR SIMPLE U-NET, NN-U-NET AND IMPROVED NNU-NET

Method	Evaluation Score			
	Dice	mIoU	Recall	Precision
U-Net	0.6669	0.5121	0.8673	0.5763
nnU-Net	0.8274	0.7366	0.8679	0.8210
Improved nnU-Net	**0.8405**	**0.7556**	**0.8840**	**0.8316**

Fig. 3 shows the segmentation results of the model on the test dataset. As can be seen from the figure, this segmentation model can accurately segment the zygomatic bone in the US images, laying the foundation for achieving high-precision 3D reconstruction and registration.

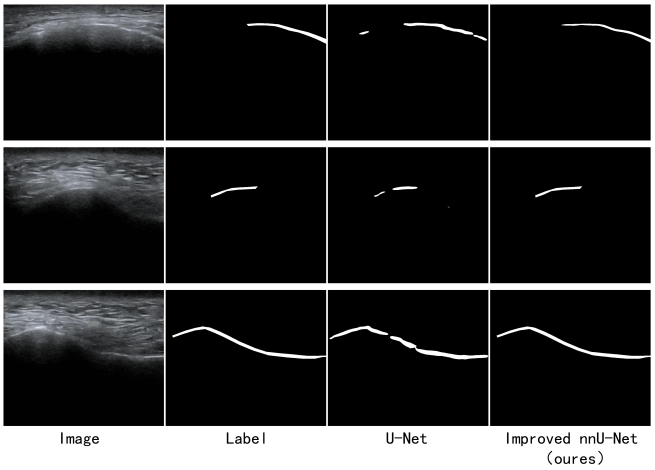

Fig. 3. Segmentation results of U-Net and the improved nnU-Net

B. 3D Reconstruction and Registration

As shown in Fig. 4, we designed two experiments to assess the accuracy of 3D US reconstruction and registration. In Experiment 1, a human volunteer was equipped with a head-mounted marker, which significantly eliminated the impact of head movement compared to not using it. US images of the volunteer's maxillofacial region were collected for reconstruction. In Experiment 2, the marker was affixed to a skull phantom, which was scanned by a surgeon in a water

tank. The phantom and the registration results in Experiment 2 were subsequently employed in robotic osteotomy trials.

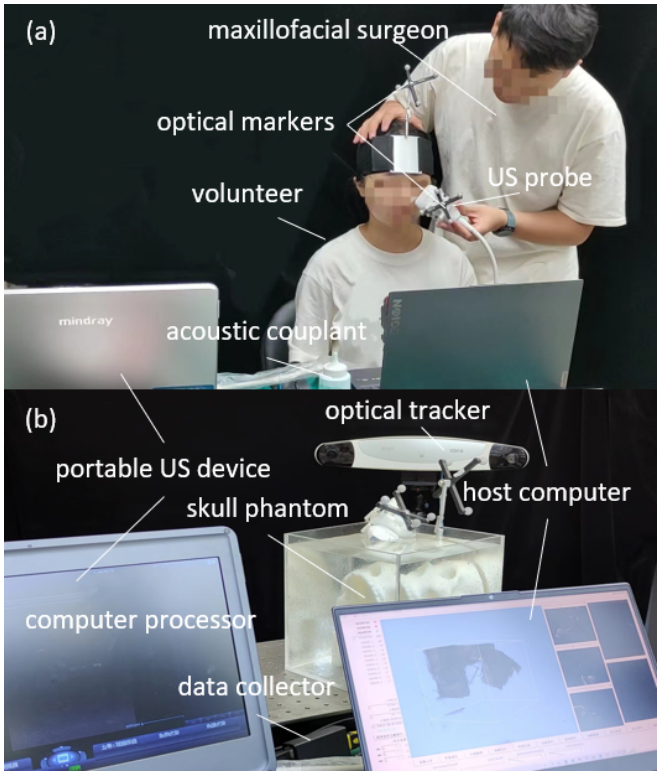

Fig. 4. 3D US reconstruction experimental scene. (a) human volunteer experiment scene. The surgeon scanned the volunteer's frontal and zygomatic bone with a US probe, an optical marker was attached to the volunteer's head (b) phantom experiment scene. The skull phantom was submerged in water to facilitate US imaging. The optical tracker was used to capture the spatial pose of the image.

Fig. 5 and Table II present the outcomes of the 3D reconstruction and registration processes. The registration error was quantified as the root-mean-square (RMS) distance between matching points in the 3D reconstruction model and the CBCT model. A total of 14 trials were conducted, and the average errors were listed. Due to the obstruction of the head-mounted marker and the non-ideal shape of human head for probe scanning, the reconstructed area of the skull in Experiment 1 was smaller compared to Experiment 2. However, in the registration stage, both 3D reconstruction models can successfully match with CBCT data, and the errors were all less than 1 mm, thereby satisfying the requirements of surgical navigation. Among them, the minimum error achieved in Experiment 1 was 0.55 mm, and in Experiment 2 was 0.51 mm.

C. Robotic Maxillary Osteotomy Experiment

Initially, under the admittance control mode, the surgeon positioned the robot near to the osteotomy site. Then, guided by the optical tracker, the robot autonomously executed the LeFort I osteotomy consistent with the preoperative plan. The experimental setup is shown in Fig. 6. An ATI six-dimensional force sensor was mounted on the robot's flange to gather force data. The end-effector in this system was

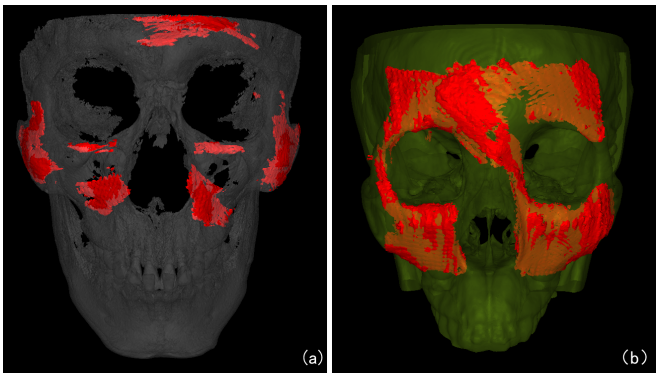

Fig. 5. 3D US reconstruction and registration results. (a) Experiment 1. (b) Experiment 2.

TABLE II
REGISTRATION RESULTS OF THE 3D RECONSTRUCTION MODEL OF HUMAN AND PHANTOM

Scanned object	Scanned area	RMS_{error}(mm)
Human	Unilateral zygomatic bone	0.79 0.65 0.55 0.60
Human	Zygomatic bone and frontal bone	0.56 0.71 0.81 0.83
Phantom 1	Zygomatic bone and frontal bone	0.67 0.68 0.69
Phantom 2	Zygomatic bone and frontal bone	0.51 0.57 0.57

a reciprocating saw specifically designed for maxillofacial osteotomy. The skull phantoms used were 3D printed resin models.

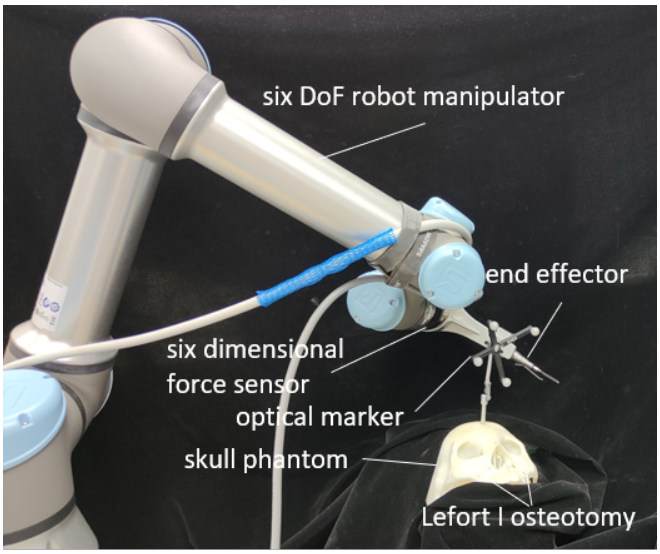

Fig. 6. Robot automatic osteotomy experiment scene.

The experiment was conducted on two skull phantoms.

After robotic osteotomy, the skull phantoms were subjected to CBCT scanning and subsequently registered with the pre-operative plan to evaluate osteotomy precision. Five points were randomly selected along the osteotomy line, from the inception to conclusion. At these points, we measured the deviation between the intended and executed osteotomy lines, calculating their average distance as the metric for the osteotomy accuracy. Considering the saw's thickness of 0.8 mm, the true measured accuracy equates to the recorded value ±0.4 mm. The results of osteotomy accuracy are illustrated in Fig. 7 and Fig. 8.

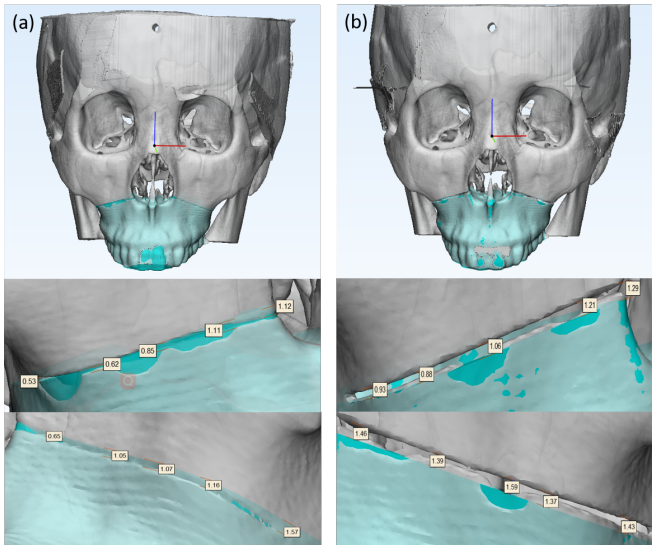

Fig. 7. LeFort I osteotomy experiments on two skull phantoms. The subsequent two lines depict magnified views of the osteotomy lines for the right and left maxilla respectively, along with measurements of the osteotomy accuracy. (a) phantom A, (b) phantom B

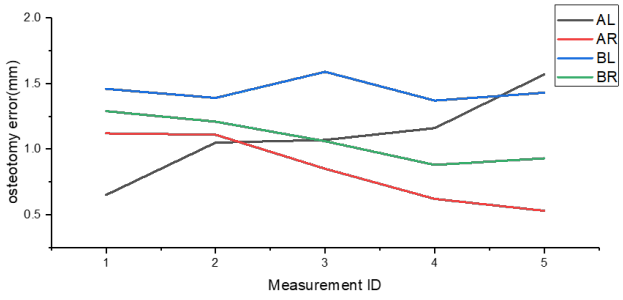

Fig. 8. Experimental data of the robotic maxillary osteotomy experiments on two skull phantoms. "AL" and "AR" denote the left and right maxilla of Model A respectively. The same for "BL", "BR".

The osteotomy accuracy depends on various factors. In addition to the accuracy of the US-based registration, it is affected by the calibration accuracy of the US probe, the accuracy of the robot's hand-eye calibration, the calibration accuracy linking the end-effector and the robot's flange, as well as the tracking accuracy of the optical tracker, among others. Table III presents a summary of the osteotomy errors, with the maximum error less than 1.5 mm.

TABLE III
ROBOTIC OSTEOTOMY ERROR (MM)

	AL	AR	BL	BR
mean	1.10	0.85	1.45	1.07
max	1.57	1.12	1.59	1.29
min	0.65	0.53	1.37	0.88
max-min	0.92	0.59	0.22	0.41

IV. DISCUSSION

In this study, we have proposed a robotic system for maxillary osteotomy, featuring intraoperative freehand 3D US reconstruction and registration. The application of US to retrieve the superficial bone surfaces for CMF surgical robot with non-invasiveness and non-radiation, is relatively novel in this context.

Our developed framework for reconstructing 2D US images, leveraging the nnU-Net neural network and DSBezier method, has demonstrated the capability to efficiently segment and reconstruct the skull's surface. Notably, it effectively filled void voxels, enabling rapid and smooth reconstructions. The registration accuracy was experimentally investigated to be less than 1 mm by comparing the shape between the 3D US reconstructions and the CT scans, confirming the method's reliability for subsequent surgical applications.

Although this method has achieved non-invasive and accurate image registration, there is still a drawback. The time required for data processing is slightly long. Each scan usually generates about 2000 US images. Due to the use of deep learning neural networks and function-based reconstruction methods, it takes several minutes to process these images. Future efforts will focus on mitigating this challenge through the adoption of more lightweight deep learning models and parallel computing.

The accuracy of the robotic osteotomy is subject to a variety of error sources, such as reconstruction accuracy, calibration accuracy, optical tracker's tracking accuracy and robot control accuracy, and the saw blade's deformation. Our phantom skull osteotomy trials yielded an average error below 1.5 mm, with the minimum error less than 1 mm. Nevertheless, additional validation involving human subjects are essential to further confirm the system's repeatability.

V. CONCLUSION

This study presents a robot-assisted system designed for CMF surgery based on intraoperative freehand 3D US reconstruction and registration. The 3D US reconstruction error for both phantoms and a human volunteer was consistently below 1 mm, aligning with the clinical requirements. The osteotomy trials on phantoms showed an average osteotomy error below 1.5 mm, showing the potential practicality in clinic. The inherent benefits of the US imaging, coupled with the accuracy demonstrated by our approach, highlight the system's promising potential. In the future, we will undertake additional experiments and enhancements to explore the clinical viability of this system.

REFERENCES

[1] F. Borumandi, M. Heliotis, C. Kerawala, B. Bisase, and L. Cascarini, "Role of robotic surgery in oral and maxillofacial, and head and neck surgery," *BRITISH JOURNAL OF ORAL & MAXILLOFACIAL SURGERY*, vol. 50, no. 5, pp. 389–393, JUL 2012.

[2] J.-H. Zhu, J. Deng, X.-J. Liu, J. Wang, Y.-X. Guo, and C.-B. Guo, "Prospects of robot-assisted mandibular reconstruction with fibula flap: Comparison with a computer-assisted navigation system and freehand technique," *JOURNAL OF RECONSTRUCTIVE MICROSURGERY*, vol. 32, no. 9, pp. 661–669, NOV 2016.

[3] L. Lin, Y. Shi, A. Tan, M. Bogari, M. Zhu, Y. Xin, H. Xu, Y. Zhang, L. Xie, and G. Chai, "Mandibular angle split osteotomy based on a novel augmented reality navigation using specialized robot-assisted arms—a feasibility study," *Journal of Cranio-Maxillofacial Surgery*, vol. 44, no. 2, pp. 215–223, 2016.

[4] J. J. Han, S.-Y. Woo, W.-J. Yi, and S. J. Hwang, "A robot arm and image-guided navigation assisted surgical system for maxillary repositioning in orthognathic surgery: A phantom skull-based trial," *Applied Sciences*, vol. 10, no. 4, p. 1549, 2020.

[5] L. Shao, X. Li, T. Fu, F. Meng, Z. Zhu, R. Zhao, M. Huo, D. Xiao, J. Fan, Y. Lin *et al.*, "Robot-assisted augmented reality surgical navigation based on optical tracking for mandibular reconstruction surgery," *Medical Physics*, 2023.

[6] Q. Ma, E. Kobayashi, H. Suenaga, K. Hara, J. Wang, K. Nakagawa, I. Sakuma, and K. Masamune, "Autonomous surgical robot with camera-based markerless navigation for oral and maxillofacial surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 2, pp. 1084–1094, 2020.

[7] H. Deng, H. Bian, C. Li, and Y. Li, "Autonomous dental robotic surgery for zygomatic implants: A two-stage technique," *The Journal of Prosthetic Dentistry*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022391323004080

[8] B. Šiniković, F.-J. Kramer, G. Swennen, H.-T. Lübbers, and R. Dempf, "Reconstruction of orbital wall defects with calcium phosphate cement: clinical and histological findings in a sheep model," *International journal of oral and maxillofacial surgery*, vol. 36, no. 1, pp. 54–61, 2007.

[9] K.-F. Hung, F. Wang, H.-W. Wang, W.-J. Zhou, W. Huang, and Y.-q. Wu, "Accuracy of a real-time surgical navigation system for the placement of quad zygomatic implants in the severe atrophic maxilla: A pilot clinical study," *Clinical Implant Dentistry and Related Research*, vol. 19, no. 3, pp. 458–465, 2017.

[10] F. Wang, M. M. Bornstein, K. Hung, S. Fan, X. Chen, W. Huang, and Y. Wu, "Application of real-time surgical navigation for zygomatic implant insertion in patients with severely atrophic maxilla," *Journal of Oral and Maxillofacial Surgery*, vol. 76, no. 1, pp. 80–87, 2018.

[11] S. Alp, M. Dujovny, M. Misra, F. Charbel, and J. Ausman, "Head registration techniques for image-guided surgery," *Neurological research*, vol. 20, no. 1, pp. 31–37, 1998.

[12] G. Grevers, A. Leunig, A. Klemens, and H. Hagedorn, "Cas of the paranasal sinuses–technology and clinical experience with the vector-vision-compact-system in 102 patients," *Laryngo-rhino-otologie*, vol. 81, no. 7, pp. 476–483, 2002.

[13] R. Marmulla, T. Lüth, J. Mühling, and S. Hassfeld, "Automated laser registration in image-guided surgery: evaluation of the correlation between laser scan resolution and navigation accuracy," *International journal of oral and maxillofacial surgery*, vol. 33, no. 7, pp. 642–648, 2004.

[14] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[15] D. G. Gobbi and T. M. Peters, "Interactive intra-operative 3d ultrasound reconstruction and visualization," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*, T. Dohi and R. Kikinis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 156–163.

[16] Q. Huang, Y. Huang, W. Hu, and X. Li, "Bezier interpolation for 3-d freehand ultrasound," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 3, pp. 385–392, 2014.

[17] R. Zhang, B. Jie, Y. He, L. Zhu, Z. Xie, Z. Liu, H. Mo, and J. Wang, "Craniomaxillofacial bone segmentation and landmark detection using semantic segmentation networks and an unbiased heatmap," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, pp. 427–437, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:265505788

[18] L. Mercier, T. Langø, F. Lindseth, and L. Collins, "A review of calibration techniques for freehand 3-d ultrasound systems," *Ultrasound in medicine & biology*, vol. 31, pp. 449–71, 05 2005.

[19] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[20] E. B. Strong, A. Rafii, B. Holhweg-Majert, S. C. Fuller, and M. C. Metzger, "Comparison of 3 Optical Navigation Systems for Computer-Aided Maxillofacial Surgery," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 134, no. 10, pp. 1080–1084, 10 2008.

[21] D. Avola, L. Cinque, A. Fagioli, G. Foresti, and A. Mecca, "Ultrasound medical imaging techniques: a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–38, 2021.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[23] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*. Springer, 2017, pp. 379–387.

[24] P. Mishra and K. Sarawadekar, "Polynomial learning rate policy with warm restart for deep neural network," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, 2019, pp. 2087–2092.

[25] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '87. New York, NY, USA: Association for Computing Machinery, 1987, p. 163–169. [Online]. Available: https://doi.org/10.1145/37401.37422