Spatially Compact Visual Navigation System for Automated Suturing Robot Towards Oral and Maxillofacial Surgery

Shaoan Wang, Qiming Zhao, Dongyue Li, Yaoqing Hu, Mingzhu Zhu, Fusong Yuan, Jinyan Shao, and Junzhi Yu, *Fellow, IEEE*

***Abstract*—The development of an automated suturing robot for oral and maxillofacial surgery (OMS) has long been a challenging task, mainly due to the spatial constraints in the oral cavity, which make intraoral navigation difficult to realize. In this paper, we propose a spatially compact visual navigation system for a single-arm suturing robot to enable real-time high-precision navigation during the intraoperative stage. Correspondingly, a two-stage intraoperative navigation framework is designed to achieve intraoral and extra-oral navigation. By designing a novel mouth opener with a fixed endoscope, the navigation information from different modules can be aligned. By strategically placing flexible position-sensitive visual markers on the robot body, end-effector, and mouth opener, precise pose information can be acquired without taking up additional space, thus guiding the robot to automatically perform the suturing process. In addition, we design a normal vector guidance-bundle adjustment (NVG-BA) module and a dynamic ROI extraction module to improve the system performance. We constructed a real suturing robot system using the proposed visual navigation framework and designed a large number of localization accuracy experiments. All of these experiments collectively demonstrate the exceptional localization accuracy and safety provided by the proposed navigation system, surpassing previously reported OMS navigation systems, all while avoiding any additional occupation of intraoral space. Finally, the system completed an automated suturing process on a simulated wound of a head phantom, demonstrating the automated suturing capability of the system.**

***Index Terms*—Visual navigation, suturing robot, flexible marker, extra-oral position navigation, intraoral pose adjustment.**

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1312800, in part by the National Natural Science Foundation of China under Grant 62373109 and Grant 82371019, in part by the Peking University Clinical Scientist Training Program BMU2023PYJH019, and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Junzhi Yu.)*

Shaoan Wang, Dongyue Li, Yaoqing Hu, Jinyan Shao, and Junzhi Yu are with the State Key Laboratory for Turbulence and Complex Systems, Department of Advanced Manufacturing and Robotics, College of Engineering, Peking University, Beijing 100871, China (e-mail: wangshaoan@stu.pku.edu.cn; 2101111894@stu.pku.edu.cn; 2001111648@stu.pku.edu.cn; jyshao@pku.edu.cn; junzhi.yu@ia.ac.cn).

Qiming Zhao is with the School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China (e-mail: MangoZhao@buaa.edu.cn).

Mingzhu Zhu is with the Department of Mechanical Engineering, Fuzhou University, Fuzhou 350000, China (e-mail: mzz@fzu.edu.cn).

Fusong Yuan is with the National Engineering Laboratory for Digital and Material Technology of Stomatology, Center of Digital Dentistry, Peking University School and Hospital of Stomatology, Beijing 100190, China (e-mail: yuanfusong@bjmu.edu.cn).

I. Introduction

FULLY automated surgical procedures are becoming a popular research topic in the field of medical engineering [1], [2]. Traditional surgical procedures are influenced to a certain extent by the fineness of the surgeon's hand, fatigue, and hand-eye coordination. However, the advent of surgical robots will ideally eliminate this limitation. Accurate and robust localization and environment perception are essential to achieve fully automated surgical procedures based on surgical robots. Intelligent surgical navigation systems are currently undergoing significant development. With different kinds of sensors (image [3], [4], infrared [5], [6], electromagnetic [7], etc.), the pose information of the robot and the patient can be acquired and used to guide the robot for fully automated surgery.

Oral and maxillofacial surgery (OMS) is widely used for the treatment of facial and oral diseases, including trauma, facial infections, and oral tumors [8]. In traditional OMS, surgeons must carefully maneuver through the narrow mouth to perform suturing procedures [9]. The constricted space and visual obstacles challenge the surgeon's experience and physical stamina, as well as limiting the surgeon's ability to observe the patient's operative region. With the rapid development of computer technology, computer-assisted surgery (CAS) has been widely used in surgical procedures to provide surgeons with a precise and powerful assistive positioning system that minimizes surgical risk, improves surgical accuracy, and improves patient prognosis [10]. Among the many navigation systems, vision-based navigation systems are gradually becoming mainstream due to their low cost, easy implementation, and high environmental tolerance.

Today's visual navigation systems face many challenges and shortcomings. First, current navigation systems are primarily placed outside the patient's body. While this design has proven effective in certain surgical procedures, it is clearly difficult to acquire the pose of the robot end-effector in the oral cavity from outside the body [11]. In addition, conventional visual navigation systems also tend to consume significant spatial resources and typically require localization fiducials to be mounted on both the patient and the robot, which is similarly inaccessible in the constricted oral cavity. In recent years, efforts have been made to utilize the patient's own features for markerless visual navigation. However, this approach may reduce localization accuracy and still fails to overcome the motion limitations of oral surgery robots.

To overcome these drawbacks and dilemmas, this paper provides a spatially compact visual navigation system for oral surgery suturing. To address the inherent spatial constraints of the oral cavity, this paper introduces a flexible, position-sensitive visual marker called HydraMarker [12], which can be tightly fitted to the robot and mouth opener without imposing additional spatial burdens. Based on this navigation system, this paper also designs a two-stage navigation framework and achieves precise intraoral localization of the robot end-effector in the oral cavity by a novel mouth opener. In addition, in order to improve the stability of localization, this paper also introduces a pose refinement method called normal vector guided-bundle adjustment (NVG-BA). Finally, we also propose a dynamic ROI extraction algorithm to improve the operational efficiency of the navigation system. In other words, the main contributions of this paper are as follows:

1) A spatially compact visual navigation system is proposed for OMS suture surgery in confined spaces, and a two-stage intraoperative autonomous navigation framework is designed for this system.
2) A pose refinement algorithm based on normal vector guidance is proposed for this visual navigation system to improve the robustness of visual localization and a dynamic ROI algorithm is also proposed to improve the operational efficiency of the visual navigation system.
3) A physical fully automated suturing robot system for OMS is developed using the proposed visual navigation system. Various experiments validate the effectiveness of the system and demonstrate its superiority over conventional surgical procedures.

The remainder of this paper is organized as follows. In Section II, we provide a comprehensive overview of the research related to existing OMS visual navigation systems and their application to suture robots. The design of the navigation system is described in detail in Section III. Then, the extensive experiments conducted in this paper are presented in Section IV, which also includes the analysis of the experimental results. Section V contains a discussion and conclusion to consider possible future work.

II. Related Works

A. Visual Navigation Systems for OMS

Visual navigation systems for OMS can be divided into two main categories: marker-based and marker-free navigation systems. Marker-based navigation systems acquire high-precision poses of the surgical tool and surgery targets by anchoring a visual marker in the oral cavity. Most of these navigation systems place a target marker on the patient's teeth and use a stereo-vision system to recognize the marker features and match them with a standard model to recover the poses. Block et al. [13] designed a cylinder-shaped mouth opener with visual markers for fixation to the patient's oral cavity, obtained a high-precision position of the patient's head by recognizing the visual markers with a binocular camera, and accomplished a teeth implant surgery. Yang et al. [14] affixed planar markers to the mouth opener and designed a vision-guided robotic system for fully automated implant surgery. Hu et al. [15] designed a surgical tool equipped with a miniature camera and replaced the markers on the mouth opener with a camera for oral localization by tracking an external checkerboard grid, resulting in a lightweight design of the navigation system. However, the operating area of the OMS is very limited, so bulky markers may become an obstacle during surgery. Furthermore, for intraoral suturing procedures, the narrow space likewise severely limits the options for placement of markers on the robot end-effector. On the contrary, marker-free navigation methods have gradually attracted the attention of researchers in recent years. This kind of visual navigation system abandons high-precision visual markers and uses a stereo vision system to extract features directly from the patient's head and align them with the 3D CT model to discard the spatial constraint problem of visual markers. Wang et al. [16], [17] applied augmented reality to surgical navigation by matching the patient's CT teeth model with features captured by a binocular camera. Ma et al. [18] obtained the patient's head position by creating a large amount of 2D contour data to build a shape-based offline model and aligning the recognized tooth contours with the model using a monocular camera. Li et al. [19] proposed an efficient texture-less pose estimation method using only teeth shape information. However, this marker-free scheme is very sensitive to noise and cannot achieve the stability and accuracy of high-precision markers. In addition, for intraoral suture surgery, its low-light and weak texture environment also makes the extraction process of naturally occurring features difficult.

B. Vision Systems on Suture Robots

Several existing fully automated suture robots have also introduced visual sensors to assist in localization. StapBot [20] utilizes an optical proximity sensor as a basis for position measurement, but as a rudimentary work does not exploit the image. Wang et al. [21] designed a suture robot with two sets of binocular cameras. The binocular camera outside the robot is used to localize the robot and the other is used to localize the wound position. However, there is a lack of specific description and accuracy analysis of the wound recognition algorithm. Li et al. [22] proposed a visual learning network for monocular needle pose estimation. However, this deep learning-based approach requires a high degree of viewpoint consistency and yields unstable pose estimations. Bendikas et al. [23] used reinforcement learning to design a framework for an external camera-based multi-stage pick-and-place needle manipulation task. Yet only the method was validated in a simulator and physical transfer experiments were lacking. Pedram et al. [24], [25] designed a vision system similar to [21], but the cameras on the robot are only used to recognize pre-drawn red dots near the wound and have limited practical application. Therefore, a visual navigation system for suturing robots still needs to be developed.

III. Spatially Compact Visual Navigation System

A. System Overview and Coordinate Formulation

The proposed visual navigation system is oriented towards a novel single-arm suturing robot [26]. This suturing robot

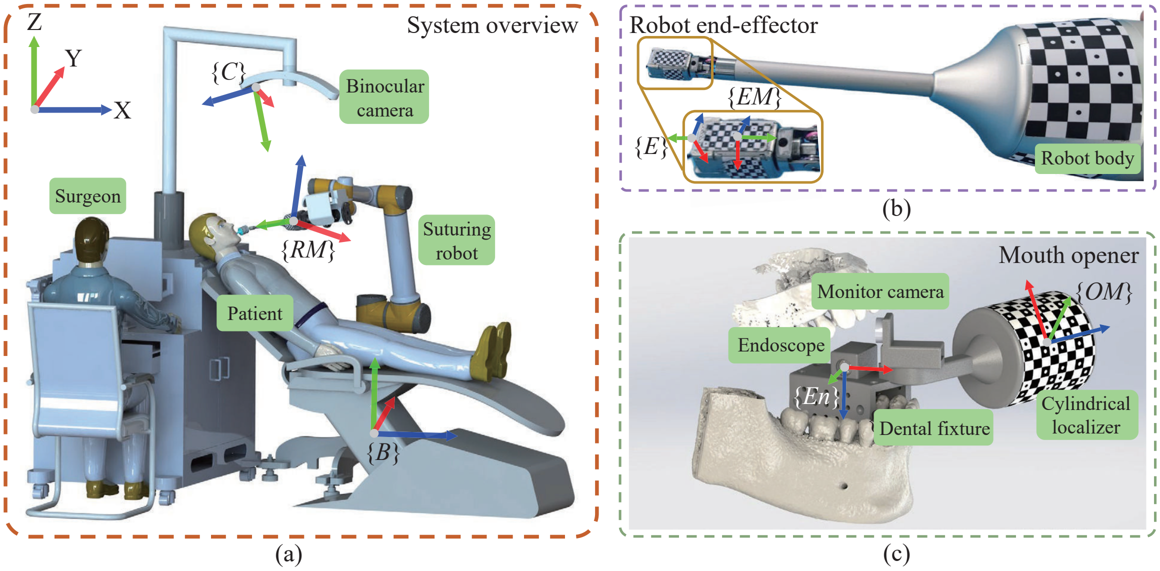


Fig. 1. Schematic diagram of the proposed navigation system, including the overview, modules, and coordinate frames.

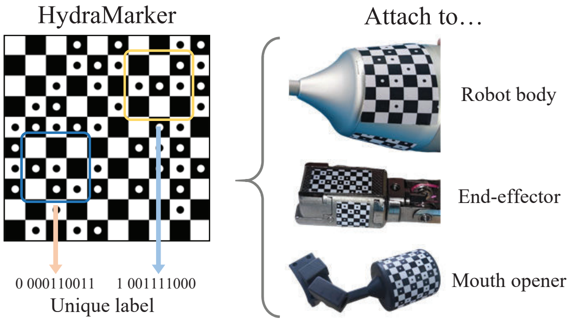


Fig. 2. The HydraMarker used in the proposed navigation system contains a unique label for each small area for reliable identification. In the proposed navigation system, multiple HydraMarkers are attached to the robot body, end-effector, and mouth opener, respectively.

employs a tendon drive mechanism with an external drive unit and utilizes only a multi-DoF end-effector to achieve flexible movement within the oral cavity. Such oral surgery robots often have tiny end-effectors to realize precise intraoral movements to adapt to the limited space inside the oral cavity, as shown in Fig. 1(b). Due to the limited space in the oral cavity and severe line-of-sight occlusion, it is difficult for conventional visual navigation systems to place large fiducials on the tiny end-effector and detect them consistently, thus making it impossible to obtain the navigation status in the oral cavity.

To solve these problems, we designed a spatially compact navigation system based on visual markers as shown in Fig. 1(a). The visual navigation system consists of the following modules: a binocular camera, a mouth opener with an endoscope, and visual markers that are attached to the robot body, end-effector, and mouth opener. The binocular camera is used to simultaneously detect visual markers on the robot body and the mouth opener, and recover their corresponding poses. As shown in Fig. 1(c), a novel mouth opener has been meticulously designed to preserve the original function while catering to the constrained space within the oral cavity, thereby enhancing the potential of intraoral navigation. Notably, this mouth opener is equipped with an endoscope, allowing for real-time visualization inside the oral cavity. This facilitates the detection of visual markers on the end-effector as it enters the oral cavity, enabling accurate estimation of its pose. Such a design offers superior insight into the intraoral environment compared to conventional surgical navigation systems, thereby fostering a navigation process characterized by precision and safety. In addition, a cylindrical localizer is provided, which serves as an intermediary between the inside and outside of the oral cavity through its recognition by the binocular camera. Finally, a monitor camera is incorporated to help the surgeon monitor the status of the operative area in real time. For the choice of visual marker, the proposed system introduces a flexible position-sensitive visual marker called HydraMarker illustrated in Fig. 2. A number of advantages exist for this visual marker. First, this marker provides a large number of features, thus providing redundant pose estimation ability. In addition, this marker possesses self-identifiability, i.e., each small region contains a unique label, which can be recognized even when a small portion of the marker is observed. Finally, it exhibits some resistance to cutting and bending, allowing it to be attached to any regular surface, as illustrated in Fig. 1(b)(c). As a result, this visual marker can adapt well to various potential contingencies, such as line-of-sight occlusion, partial contamination of the marker, and illumination changes, thereby greatly enhancing the robustness and the long-term reliability of the visual navigation system. The recognition algorithm and the 3D reconstruction algorithm for this visual marker can be found in [15], [27], respectively.

Fig. 1 illustrates specifically the coordinate system definition and implementation process of the proposed visual navigation system. We utilize $\bullet_i^j$ to describe the information of frame i as presented in frame j. Meanwhile, $p_i^j \in \mathbb{R}^3$ and $P_i^j \in \mathbb{R}^4$

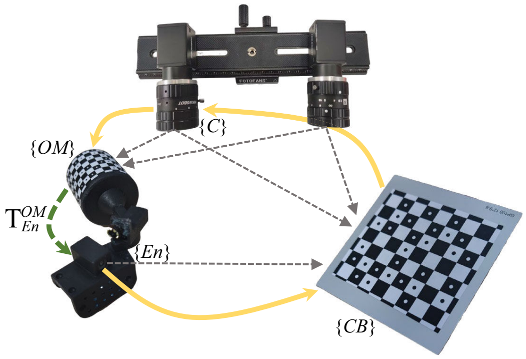


Fig. 3. Illustration of the calibration between the mouth opener and the endoscope.

represent points of frame i related to frame j in Cartesian and Homogeneous, respectively. There are a total of eight frames defined to address the proposed navigation system (see Fig. 1): robot base $\{B\}$, camera $\{C\}$, endoscope $\{En\}$, opener marker $\{OM\}$, end-effector $\{E\}$, robot marker $\{RM\}$, end-effector marker $\{EM\}$, and end-effector zero position $\{E_0\}$.

B. Calibration of Navigation System

Upon the initial deployment of the proposed navigation system, calibration becomes imperative to refine the transformation relationships among diverse sources of navigation information. It merits emphasis that in a stable operational setting, calibration is a one-time procedure. Once calibration is successfully executed, the system becomes ready for recurrent utilization. This subsection will provide a complete calibration solution for the proposed navigation system.

1) Calibration between mouth opener with endoscope: During the intraoral navigation process, this transformation pathway from the endoscope to the robot base is required to obtain an accurate endoscope-to-opener transformation relationship. We derive the transformation matrix $\mathbf{T}^{OM}_{En}$ by using a properly sized visual marker (here, for consistency in the recognition algorithm, we use a checkerboard with the same HydraMarker) $\{CB\}$, which is placed in the common field of view of the endoscope and binocular camera as shown in Fig. 3. The introduction of the visual marker offers a transformation relationship between the endoscope with the mouth opener:

$$\mathbf{T}^{C}_{OM}\mathbf{T}^{OM}_{En}\mathbf{T}^{En}_{CB} = \mathbf{T}^{C}_{CB} \tag{1}$$

Therefore, $\mathbf{T}^{OM}_{En}$ can be obtained from the following equation:

$$\mathbf{T}^{OM}_{En} = (\mathbf{T}^{C}_{OM})^{-1}\mathbf{T}^{C}_{CB}(\mathbf{T}^{En}_{CB})^{-1} \tag{2}$$

It is worth noting that both poses $\mathbf{T}^{C}_{OM}$ and $\mathbf{T}^{En}_{CB}$ acquired by this navigation system are valid, non-singular configurations. As a result, each pose corresponds to a unique solution for the corresponding $\mathbf{T}^{OM}_{En}$.

2) Calibration of suturing robot: At this point, there are still two fixed but unknown quantities, $p^{RM}_{E_0}$ and $\mathbf{R}^{RM}_{E_0}$. As illustrated in Fig. 4, we use a surgical instrument with a calibrated pivot [28] to calibrate $p^{RM}_{E_0}$ by touching the surgical instrument pivot to the origin of the end-effector, i.e. $p^{C}_{SI} = p^{C}_{E_0}$, therefore:

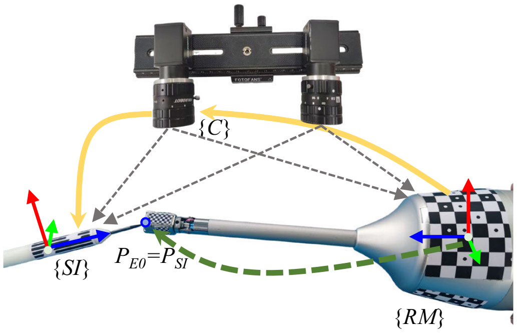


Fig. 4. Schematic diagram of the end-effector calibration of the suturing robot.

$$P^{RM}_{E_0} = (\mathbf{T}^{C}_{RM})^{-1}P^{C}_{E_0} = (\mathbf{T}^{C}_{RM})^{-1}P^{C}_{SI} \tag{3}$$

Due to the rigid body assumption, we calibrate $\mathbf{R}^{RM}_{E_0}$ using robot arm kinematics:

$$\mathbf{R}^{RM}_{E_0} = (\mathbf{R}^{C}_{RM})^{-1}(\mathbf{R}^{B}_{C})^{-1}\mathbf{R}^{B}_{J_6} \tag{4}$$

C. Two-Stage Intraoperative Navigation Framework

Fig. 5 depicts the framework of the navigation system, illustrating its division into two main components: offline preliminaries and online intraoperative navigation. Our visual navigation system is designed around a two-stage intraoperative navigation framework. This framework delineates the intraoperative navigation process into two distinct stages: the extraoral position navigation stage and the intraoral pose adjustment stage. In the extraoral position navigation stage, the system primarily identifies markers on the robot body and mouth opener to ascertain their relative poses. This information is then utilized to control the robot arms movement towards the target position. Conversely, the intraoral pose adjustment stage involves recognizing markers on the robot end-effector via the endoscope. This enables the system to determine the relative poses between the robot body and the end-effector, facilitating precise control of the end-effector to achieve specified poses. It is paramount to note that marker recognition in each frame during intraoperative navigation occurs independently, rendering the corresponding pose unaffected by previous results. This subsection concretely describes the implementation details of this framework.

1) Extra-oral position navigation process: The main purpose of the extra-oral navigation process is to guide the suture end-effector into the oral cavity, and subsequently to guide the end-effector to the suture position after the intraoral pose adjustment has been completed. This process assumes that the robot end-effector remains fixed to the robot body, i.e., it does not change the pose of the end-effector with respect to the robot arm joint 6 $\{J_6\}$. Thus, the positional relation of $\{E_0\}$ with respect to $\{B\}$ is as follows:

$$\mathbf{T}^{B}_{E_0} = \mathbf{T}^{B}_{C}\left[\begin{array}{c:c} \mathbf{R}^{C}_{RM}\mathbf{R}^{RM}_{E_0} & \mathbf{R}^{C}_{RM}p^{RM}_{E_0} + \mathbf{t}^{C}_{RM} \\ \mathbf{0} & 1 \end{array}\right] \tag{5}$$

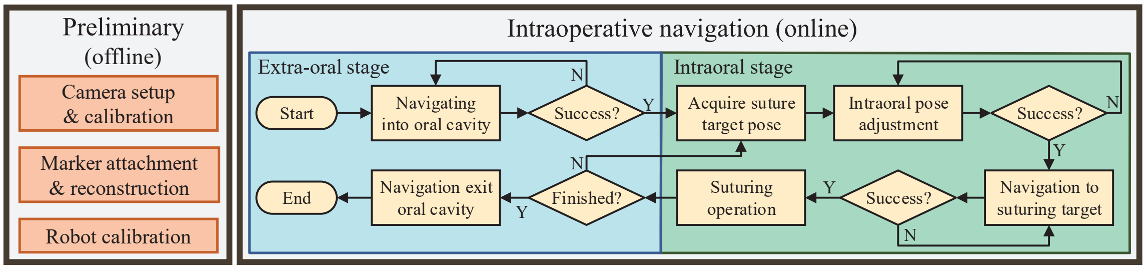

Preliminary (offline)
Camera setup & calibration
Marker attachment & reconstruction
Robot calibration
Intraoperative navigation (online)
Extra-oral stage
Start
Navigating into oral cavity
Success?
N
Y
Navigation exit oral cavity
Finished?
End
Intraoral stage
Acquire suture target pose
Intraoral pose adjustment
Success?
Navigation to suturing target
Success?
Suturing operation

Fig. 5. The flowchart of the two-stage intraoperative navigation framework.

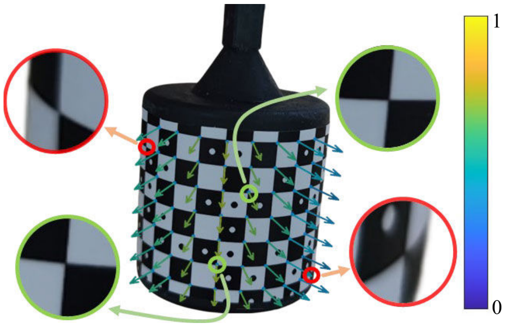


Fig. 6. Illustration of applying NVG-BA to the mouth opener. The red boxes indicate the crosspoints with severe warp, and the green boxes indicate the crosspoints with weaker warp. The vector on each crosspoint indicates its corresponding plane normal vector and the color of the vector indicates its weight.

2) Intraoral pose adjustment process: The end-effector inside the mouth can acquire its pose in real-time by endoscope recognition of the visual markers attached to it, thus enabling precise pose adjustment. Noting that the endoscope is fixedly attached to the mouth opener, therefore, the pose relation of $\{E\}$ with respect to $\{B\}$ is listed:

$$\mathbf{T}_E^B = \mathbf{T}_C^B \mathbf{T}_{OM}^C \mathbf{T}_{En}^{OM} \mathbf{T}_{EM}^{En} \mathbf{T}_E^{EM} \tag{6}$$

Similar to extra-oral position navigation, the transformation between the end-effector marker coordinate and the end-effector coordinate needs to be calibrated as well. The transformation relationship $\mathbf{T}_{EM}^E$ can be obtained using the position of the end-effector at zero position:

$$\mathbf{T}_E^{EM} = (\mathbf{T}_C^B \mathbf{T}_{OM}^C \mathbf{T}_{En}^{OM} \mathbf{T}_{EM}^{En})^{-1} \mathbf{T}_{E_0}^B \tag{7}$$

Tracking of the suture pose $\mathbf{T}_{tar}$ is achieved by the binocular camera by recognizing the visual marker on the mouth opener. Since the opener is solidly attached to the patient's head, the surgeon registers the suture positions with the mouth opener coordinate during the preoperative planning phase, thus eliminating suturing offsets caused by the patient's potential respiration or movement. During the surgery, the suture poses under the robot base $\{B\}$ satisfied:

$$\mathbf{T}_{tar_i}^B = \mathbf{T}_C^B \mathbf{T}_{OM}^C \mathbf{T}_{tar_i}^{OM} \tag{8}$$

D. Normal Vector Guided-Bundle Adjustment

For the sake of safety and stability, it is imperative that the visual navigation system of the surgical robot exhibits exceptional robustness. This means that the jitter in the target pose, caused by noise, should be minimized to ensure smooth robot motion. Traditional visual localization methods initially obtain the coarse pose by aligning the camera-detected features with their corresponding 3D spatial model [29], followed by further optimization using bundle adjustment (BA) to recover the refined pose.

During practical usage, we found that the positioning accuracy of the features has a severe influence on the optimization result. Specifically, the crosspoint features used in this paper, as illustrated in Fig. 6, may become distorted under severe warping conditions. As mentioned previously [30], this distortion often leads to a decrease in the positioning accuracy of the crosspoint, resulting in deviations in the optimized pose and undesirable jitter. To address this challenge, we introduce the Normal Vector Guided-Bundle Adjustment (NVG-BA) in this paper.

The core concept of NVG-BA is to align the normal vector direction of each corner by leveraging the distribution of feature points within the 3D model under an initial coarse pose. This approach enables the system to assess the degree of deformation experienced by each corner point relative to the current viewing angle. Consequently, different weights are assigned to each corner based on their respective degree of deformation, facilitating optimization.

For a corner $\mathbf{x}_{i,j}$ detected by the ith camera, the computation of the normal vector can be reframed as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{m},\mathbf{n}} \mathcal{J}(\mathbf{m},\mathbf{n}) &= \sum_{k\in\mathcal{N}(j)} ||\left(\mathbf{p}_{i,j,k} - \mathbf{m}\right)^T \mathbf{n}||_2 \\ \text{s.t. } \|\mathbf{n}\| &= 1 \end{aligned} \tag{9}$$

where $\mathbf{p}_{i,j,k}$ is the spatial point of the model corresponds to $\mathbf{x}_{i,j}$ and $\mathcal{N}(j)$ is the neighboring crosspoint cluster of $\mathbf{p}_{i,j,k}$. Since $\mathbf{m}$ and $\mathbf{n}$ are independent of each other, the optimization problem can be decoupled. The optimization problem is converted to when only $\mathbf{m}$ is considered:

$$arg \min_{\mathbf{m}} = \sum_{i=1}^{n} \parallel \mathbf{p}_{i,j,k} - \mathbf{m} \parallel^2 \tag{10}$$

It is easy to prove that $\mathbf{m}$ should be the center of $\mathcal{N}(j)$, hence

$$\mathbf{m} = \frac{1}{card(\mathcal{N}(j))} \sum_{k\in\mathcal{N}(j)} \mathbf{p}_{i,j,k} \tag{11}$$

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2024.3427843

let $\mathbf{q}_{i,j,k} = \mathbf{p}_{i,j,k} - \mathbf{m}$, then the optimization problem can be rewritten as:

$$\min_{\mathbf{n}} \sum_{i \in \mathcal{N}(j)} ||\mathbf{q}_{i,j,k}^T \mathbf{n}||_2 = \min_{\mathbf{n}} \mathbf{n}^T \left(\mathbf{Q}\mathbf{Q}^T\right) \mathbf{n} \qquad (12)$$
$$\text{s.t. } \mathbf{n}^T\mathbf{n} = \mathbf{1}$$

where

$$\mathbf{Q} = \begin{pmatrix} | & | & & | \\ \mathbf{q}_{i,j,1} & \mathbf{q}_{i,j,2} & \ldots & \mathbf{q}_{i,j,n} \\ | & | & & | \end{pmatrix} \qquad (13)$$

$\mathbf{n}$ is the normalized eigenvector $\mathbf{v}_{\min}$ of $\mathbf{Q}\mathbf{Q}^T$ with the smallest eigenvalue $\lambda_{\min}$ (see Appendix A for proof). Here, we define the weight of each crosspoint as being inversely proportional to the angular difference between its normal vector and the Z-axis of each camera coordinate, i.e., the more perpendicular the crosspoint is to the camera plane, the larger its corresponding weight.

$$\omega_{i,j} = \arccos\left(\mathbf{n}_{i,j} \cdot \left[\begin{array}{ccc} 0 & 0 & 1 \end{array}\right]^T\right) \qquad (14)$$

Similar to the original bundle adjustment method, we formulated the localization problem as the optimization of a cost function that contains weighted reprojection errors corresponding to the observations from the binocular camera.

$$\min J = \sum_{i=0}^{1} \sum_{j \in \mathcal{J}(i)} \mathbf{e}^{i,j^T} \mathbf{W}_r^{i,j} \mathbf{e}^{i,j} \qquad (15)$$

where i denotes the camera index and j denotes the crosspoint index. The set $\mathcal{J}(i)$ contains the indices of crosspoint detected by the ith camera. The reprojection error is:

$$\mathbf{e}_r^{i,j} = \mathbf{z}^{i,j} - \pi_i\left(\mathbf{T}_{C_iS}\mathbf{T}_{SW}\mathbf{l}^{i,j}\right) \qquad (16)$$

where $\mathbf{z}^{i,j}$ is the measured image coordinate of the jth crosspoint on the ith camera and $\mathbf{l}^{i,j}$ is the homogenerous coordinate of the jth crosspoint in the reconstruction model coordinate. Additionally, $\mathbf{W}_r^i$ is the weight matrix of the crosspoint measurement $\mathbf{l}^{i,j}$.:

$$\mathbf{W}_r^i = \operatorname{diag}\left(\omega_{i,1}, \omega_{i,2}, \cdots, \omega_{i,n}\right) \qquad (17)$$

We employ the Google Ceres optimizer to carry out the optimization procedure.

E. Dynamic ROI Extraction

The efficiency of the visual navigation system is critical to surgical performance. A high-speed navigation system improves the pose-tracking ability of the robot and suture targets and enhances the surgeon's decision-making ability to respond more quickly to unexpected situations and emergencies. To enhance the efficiency of the visual navigation system, a dynamic ROI extraction algorithm is proposed to accelerate the image processing oriented to the surgical environment of the proposed system.

In typical surgical scenarios, the quantity and arrangement of visual markers are predefined, and real-time pose information of each visual marker model is available, offering guidance for anticipating the future positions of these markers. Considering that the time interval between two consecutive frames is extremely short (generally within tens of milliseconds), the displacement of visual markers is minimal. Hence, the positions of visual markers in two adjacent frames are expected to be in close proximity. Here, we introduce a dynamic ROI extraction method. Its objective is to project the corresponding 3D model onto the image plane using the acquired positions of visual markers in the current frame. This projection enables us to predict the smallest rectangular region within which the corresponding visual marker will appear in the subsequent frame, as illustrated in Fig. 7. Consequently, only this extracted ROI region is processed in the next frame, resulting in a significant acceleration of the visual navigation system.

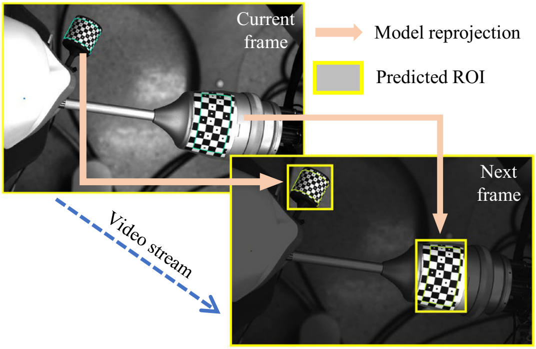


Fig. 7. Schematic diagram of the dynamic ROI extraction process.

With the marker detected in the current frame, it becomes feasible to recover the pose associated with that marker. As a result, the position of this marker will appear in the next frame can be forecast based on the camera projection model:

$$\begin{bmatrix} u_{ij}^k \\ v_{ij}^k \\ 1 \end{bmatrix} = K_i \left[\mathbf{R}_i^0 \left(\mathbf{R}_c^j P_j^k + \mathbf{t}_c^j\right) + \mathbf{t}_i^0\right] \qquad (18)$$

where i denotes the camera index, j denotes the object (robot, mouth opener, end-effector, etc.) index, and k denotes the crosspoint index. u_{ij}^k and v_{ij}^k correspond to the predicted x and y coordinates of the crosspoint in the image, respectively. K_i refers to the intrinsic parameters of camera i, and $\mathbf{R}_i^0$ refers to its extrinsic parameters with respect to camera 0. Therefore, the ROI of the jth object in camera i predicted in the next frame is:

$$\begin{cases} \left(x^{-}{}_{ij}, y^{-}{}_{ij}\right) = \left(\lceil \sup U_{ij} \rceil, \lceil \sup V_{ij} \rceil\right) \\ \left(x_{-ij}, y_{-ij}\right) = \left(\lfloor \inf U_{ij} \rfloor, \lfloor \inf V_{ij} \rfloor\right) \end{cases} \qquad (19)$$

where U_{ij} and V_{ij} represent the sets of x and y coordinates of the crosspoints of the jth object in camera i, respectively. Meanwhile, (x_{-ij}, y_{-ij}) and $(x^{-}{}_{ij}, y^{-}{}_{ij})$ denote the lower-left and upper-right coordinates of the ROI, respectively.

IV. Experiments

A. System Setup

Fig. 8 shows the physical setup of the proposed system. This system is composed of five components: the suturing robot

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2024.3427843

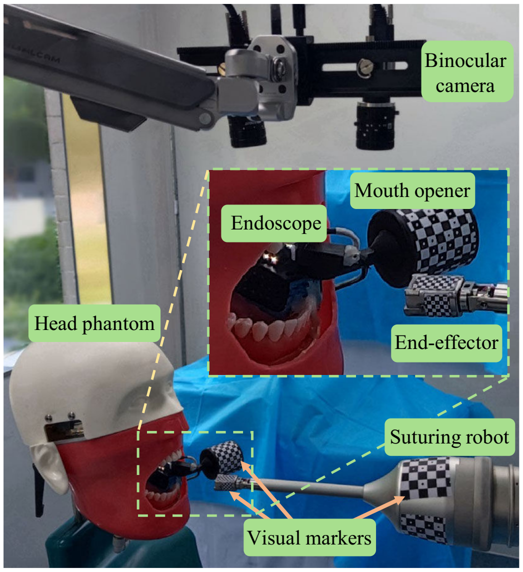


Fig. 8. The proposed spatially compact visual navigation suturing robot system.

with fiducial markers on it, a head phantom with silicone pads inside for wound simulation, a mouth opener with a fixed endoscope, the adjustable stereo vision cameras, and the computer. The suturing robot is mounted on an AUBO i5 collaborative robotic arm. The stereo vision camera used is HikVision MV-CA023-10GM industrial camera (monochrome, 1920 × 1200 pixels). The mouth opener is a 3D-printed model with a customized endoscopy (monochrome, 1920×1080 pixels). The HydraMarker and CylinderTag used on the robot and opener are printed on a PVC sticker which is less affected by ink bleeding and blood spatter. The computer is equipped with Intel I7-11700K (2.50 GHz, 32 GB RAM).

B. Suture Target Jitter Evaluation

The automated surgical process depends on acquiring the pose information of both the targets and the surgical robot, which is provided by the visual navigation system. Control and planning algorithms are subsequently employed to navigate the robot to its designated targets. Therefore, the quality of the target information provided by the visual navigation system will directly determine the control stability of the surgical robot. Most of the time, the patient will enter a certain degree of anesthesia, when the surgical targets can be considered to remain motionless. In this case, the target poses provided by the navigation system should be as stable as possible to enhance the surgical performance and reduce the risk. For this scenario, we design a jitter experiment with stationary suture targets. A mouth opener is fixedly attached to the head phantom through an impression and the phantom is kept stationary. The visual navigation system is used to continuously acquire the poses of the suture targets in the coordinates of the mouth opener and analyze their translation and rotation jitter over a period of time.

TABLE I
COMPARATIVE RESULTS OF SUTURE TARGET LOCALIZATION JITTER EXPERIMENTS IN DIFFERENT POSES

Unit: mm/°

Pose	Metric		NVG-BA	BA
"Good" pose	STD	X	**0.096**	0.108
		Y	**0.318**	0.353
		Z	**0.359**	0.377
		Rot	**0.07**	0.076
	MAD	X	**0.056**	0.065
		Y	**0.197**	0.213
		Z	**0.22**	0.225
		Rot	**0.041**	0.045
"Bad" pose	STD	X	**3.263**	3.855
		Y	**1.319**	2.203
		Z	**2.882**	3.998
		Rot	**0.19**	0.394
	MAD	X	**2.14**	3.309
		Y	**0.983**	1.761
		Z	**2.027**	3.442
		Rot	**0.141**	0.279

Due to the cylindrical shape of the mouth opener, the visual marker attached to it will inevitably be affected by the deformation, resulting in a decrease in the localization accuracy of the corners near the edge with large deformation. Therefore, there are differences in the performance of visual localization when the mouth opener is in different poses. For most "good" poses, features on the visual markers are consistently recognized and maintain high localization accuracy, which is reflected in minimal pose jitter. For some "bad" poses, features on the boundaries of the visual markers are difficult to recognize or have low localization accuracy, resulting in severe jitter in the localization results. Here, we selected a "good" and a "bad" typical pose for testing.

The degree of jitter in the NVG-BA proposed in this paper is compared with the conventional BA by calculating the standard deviation (STD) and median absolute deviation (MAD) of translation (X, Y, Z) and rotation (Rot) which are defined as respectively:

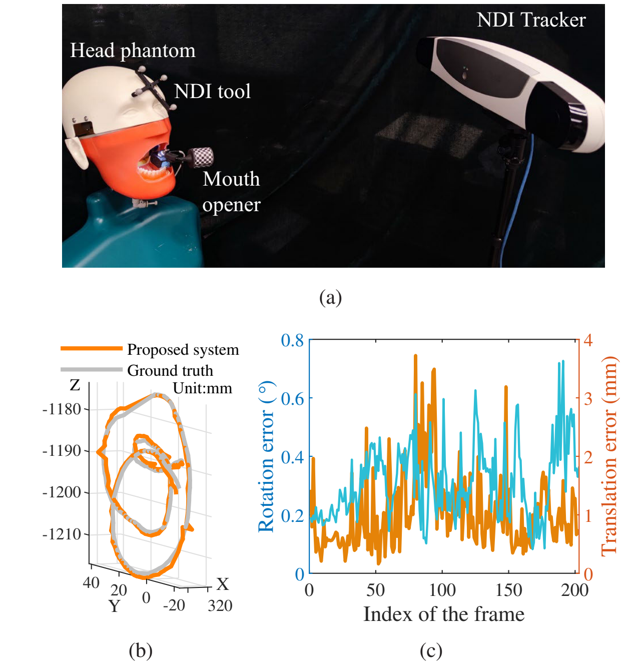
$$\begin{aligned} \mathrm{STD}(\mathbf{x}) &= \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})^2}{n-1}} \\ \mathrm{MAD}(\mathbf{x}) &= \mathrm{median}(|\mathbf{x}_i-\mathrm{median}(\mathbf{x})|) \end{aligned} \tag{20}$$

where $\mathbf{x}$ is the rotation vector or translation vector, corresponding to rotation and translation, respectively.

Table I shows the experimental results. In the case of "good" pose, both algorithms are able to maintain low-range jitter, but the NVG-BA still maintains lower jitter; while in the case of "bad" pose, the jitter suppression ability of the NVG-BA is significantly improved compared with that of the original BA, which indicates that the NVG-BA can help the navigation system to improve its stability.

C. Head Movement Localization Accuracy

Since OMS surgeries are often performed with local anesthesia, there may be uncontrollable and sudden movements of the patient's head during the procedure. Therefore, a robust navigation system should have the ability to localize head movements and provide timely and accurate feedback on the

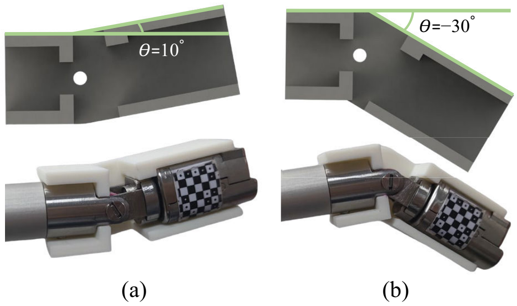


(a)

(b)

(c)

Fig. 9. Experiment setup and results of head movement localization accuracy. (a) Setup for the head movement localization experiment. (b) Head phantom trajectories measured by proposed navigation system versus ground truth trajectories. (c) Translation and rotation errors of the measured head phantom poses.



(a)

(b)

Fig. 10. Illustration of assistive gadgets mounted on the suturing robot end-effector. (a) and (b) correspond to the scenarios with tilt angle $\theta = 10°$ and $\theta = –30°$, respectively.

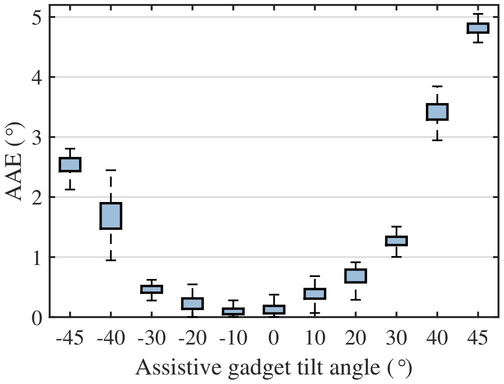


Fig. 11. Results of intraoral localization accuracy experiments.

current target position and pose. Here, we use an NDI tracker (Polaris Vega tracker, Northern Digital) to obtain ground truth pose information. The experimental setup for head movement localization accuracy is shown in Fig. 9(a). By fixing the NDI tool with passive spheres on the head phantom, its pose can be acquired using the NDI tracker. Then, the transformation relationship between the NDI tracker and the binocular camera is obtained through hand-eye calibration to align the poses acquired by them separately. During the experiment, the head poses measured by the navigation system and the NDI tracker, respectively, were recorded simultaneously. 200 consecutive frames were recorded and the translation and rotation errors between the pose estimated by the navigation system and the pose measured by the NDI were calculated.

Fig. 9(b) shows a plot of the head phantom trajectory obtained via the navigation system against the reference trajectory obtained by the NDI tracker, and Fig. 9(c) shows the translation and rotation errors for the corresponding estimated pose in each frame. The average pose estimation error for translation is 1.56 mm with a standard deviation of 0.63 mm, while the average pose estimation error for rotation is 0.21° with a standard deviation of 0.13°. The experimental results affirm that the proposed navigation system maintains a high degree of localization accuracy even in the presence of head movement, thereby confirming its safety and reliability.

D. Intraoral Localization Accuracy

The accuracy of intraoral localization determines the pose accuracy of the end-effector of a suture robot and affects the performance of the suture. To evaluate the intraoral localization accuracy of the proposed navigation system, we designed a series of 3D-printed auxiliary gadgets to fix the end-effector's pose. These auxiliary gadgets have a known tilt angle that is reflected in the yaw angle of the end-effector, as shown in Fig. 10. By simultaneously using a binocular camera to recognize visual markers on the mouth opener and an endoscope to recognize visual markers on the end-effector, the relative pose of the end-effector with respect to its zero-position pose can be obtained. Thereby, the yaw angle of the end-effector corresponding to the current pose can be calculated using the inverse kinematics of the suturing robot. We sequentially mounted the assistive gadgets on the robot end-effector and acquired 100 consecutive frames to calculate the corresponding end-effector yaw angle in the current pose. The results were compared with the corresponding tilt angles of the assist gadgets to evaluate the accuracy and stability of intraoral localization. Here, we define the Absolute Angular Error (AAE) as the experimental evaluation metric, which means the absolute value of the difference between the calculated yaw angle and the ground truth angle.

Fig. 11 illustrates the results of this experiment. Since the intraoral position accuracy is mainly determined by the selection accuracy of the end-effector origin, the further away from the zero position should theoretically cause a larger error. The corresponding experimental results also clearly show this result. It is noted that in the inclination range of -30° to 30° (which is also the most frequently used position range), our navigation system can obtain an angle resolution error of no more than 1.5°, which is quite sufficient for suture surgery.

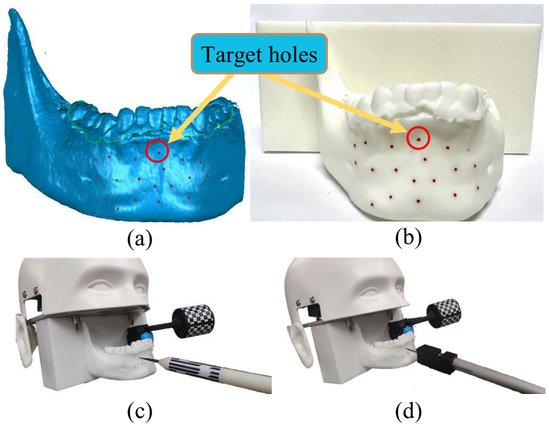


Fig. 12. Schematic of the overall navigation system accuracy experiment. (a) Fabricated CT model with target circular holes; (b) Corresponding 3D-printed model; (c) Target point registration using the surgical instrument; (d) The suturing robot end sequentially placed at the corresponding target hole.

TABLE II
COMPARATIVE RESULTS OF OVERALL VISUAL NAVIGATION ACCURACY

Methods	Ours	[16]	[19]	[17]	[15]
TRE (mm)	**0.69**	0.71	0.74	0.85	0.92

While there is an error of no more than 5° at ±40° and ±45°, it is very difficult to encounter situations that require such a large angle for a typical suturing procedure.

E. Overall Navigation System Accuracy

The overall localization accuracy of the navigation system is usually jointly affected by multiple factors that are coupled with each other. In this paper, we designed a navigation accuracy experiment for the suturing robot to evaluate the overall localization accuracy as illustrated in Fig. 12. In this experiment, we printed a CT model of the patients mandible and set a series of circular holes on its surface as target points by Boolean operations using Blender (Ver. 3.1, Blender.org). These N circular holes are in sequence registered as target points in the mouth opener coordinate using a calibrated high-precision surgical instrument [28]. Then, the suturing robot is controlled to reach this series of target points one by one. The robot end-point position information was sequentially recorded using the visual navigation system so that the distance between the registered target points and the end-points from the navigation system is utilized as the evaluation metric. Here, the target registration error (TRE) is introduced in this experiment:

$$TRE = \sqrt{\frac{\sum_{i=1}^{N} \|\boldsymbol{p}_i^c - \boldsymbol{p}_i^t\|_2^2}{N}} \tag{21}$$

where $\boldsymbol{p}_i^c$ is the ith target point registered by the surgical instrument, and $\boldsymbol{p}_i^t$ is the corresponding point detected by the navigation system. The TRE of the proposed navigation system is 0.69 mm. Compared with other methods in the open literature, our system achieves the lowest TRE as Table II shows.

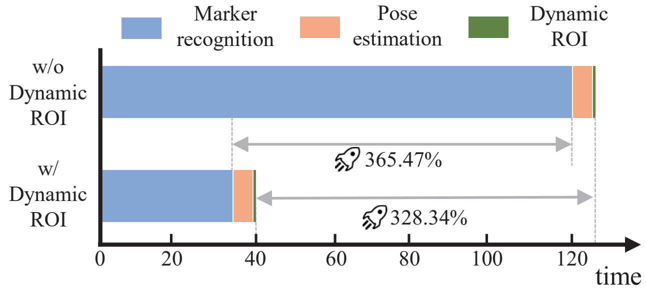


Fig. 13. The flowchart of the proposed navigation system framework.

F. Navigation System Operational Efficiency

To evaluate the computational load of the navigation system, we conducted an operational efficiency experiment. Initially, 100 sets of images were continuously captured in a simulated typical surgical scenario, with each set containing images from both views of the binocular camera and from the endoscope. Subsequently, the processing time for each image in each session (marker recognition, pose estimation, and dynamic ROI) was calculated using the proposed navigation system, and the resulting processing time for each session was averaged to determine the final operational efficiency.

It is important to note that endoscopic images are smaller in size compared to binocular camera images, resulting in shorter processing times for endoscopic images. Hence, we utilize the processing time of the binocular camera as an evaluation metric to ensure a consistent and comprehensive assessment of system efficiency.

As illustrated in Fig. 13, the experimental results reveal that the proposed navigation system operates inefficiently without the dynamic ROI module, with a processing time of 130.45 ms per frame (124.77 ms for marker recognition, 5.05 ms for pose estimation, and 0.63 ms for dynamic ROI). Notably, marker recognition accounts for the majority of the total time.

Upon integration of the dynamic ROI module, the processing time for marker recognition decreases to 34.14 ms, attributed to the increased speed of the marker recognition process by 365.47% through cropping out most irrelevant regions in the image. Consequently, the total processing time per frame reduces to 39.73 ms (34.14 ms for marker recognition, 4.92 ms for attitude estimation, and 0.67 ms for dynamic ROI), marking a significant improvement of 328.34% compared to previous results.

Thus, the proposed navigation system can operate at speeds up to 25.17 FPS, ensuring excellent real-time performance. Notably, many published papers do not provide an in-depth analysis of operational efficiency. Therefore, this paper only compares with those papers that evaluate operational efficiency. As depicted in Table III, the proposed navigation system demonstrates sufficient operational efficiency compared to conventional navigation systems, thereby ensuring the safety of autonomous surgery.

G. Automated Suture Procedure Verification

The visual navigation system proposed in this paper is deployed on a real single-arm suturing robot described in

TABLE III
COMPARATIVE RESULTS OF NAVIGATION SYSTEM OPERATIONAL EFFICIENCY

Methods	Ours	[31]	[17]	[32]
FPS	25.17	12	25	30

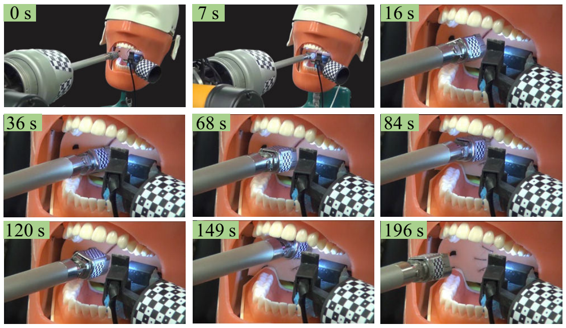
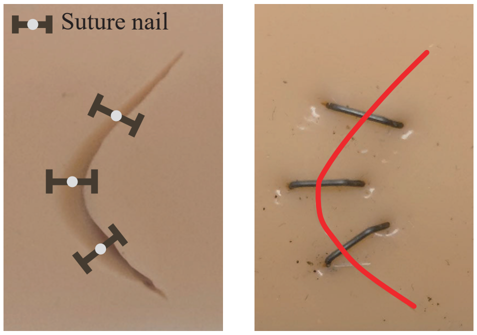


Fig. 14. The process of completing fully automated suturing of a head phantom simulated wound using the proposed navigation system.



(a) (b)

Fig. 15. Results of automated suturing procedures. (a) Schematic of the simulated wound with the results of the surgeon's preoperative planning; (b) Schematic of the wound after automated suturing operation using the proposed navigation system.

the previous section and the automatic suturing capability of this system is verified. Fig. 8 illustrates the components and arrangement of the entire system. In this validation experiment, a silicone pad with a simulated wound was placed in the head phantom to simulate an intraoral wound. The visual navigation system provides the required pose information in real time to assist the suture robot in completing the suturing task autonomously. During the entire suturing procedure, the surgeon only needs to select the current suture target in the self-developed software interface and monitor the emergency situation in the oral cavity in real time to complete the surgical process. Fig. 14 demonstrates the specific automated suturing process. Through this process, the simulated wound on the silicone pad was well sutured, as shown in Fig. 15, which verifies the effectiveness and autonomy of this system.

V. CONCLUSIONS

Fully automated intraoral suturing surgery based on visual navigation is a promising field. In this paper, a spatially compact visual navigation system for a novel single-arm suturing robot is presented. Flexible visual markers for localization are introduced to address the constraints of the narrow intraoral environment. For the proposed navigation system, a two-stage visual navigation framework is designed to synchronize the acquisition of high-precision navigation information inside and outside the oral cavity, and a complete process of system calibration and navigation information transformation is provided. We comprehensively describe the theoretical basis of these processes in this paper. Besides, to improve the localization accuracy and operational efficiency of the system, we propose a pose refinement algorithm called normal vector guided-bundle adjustment (NVG-BA) and a dynamic ROI extraction algorithm. In addition, we show a large number of experimental evaluation results for the localization accuracy of this navigation system, which all demonstrate that this scheme achieves the highest navigation accuracy without introducing additional space requirements. Finally, we applied the navigation system to a suturing robot and successfully performed a simulated automated suturing procedure on a head phantom. Together, these experiments demonstrate the superior accuracy of the visual navigation system and its efficacy in guiding the robot through automated suturing procedures.

In the future, our research efforts will continue to focus on exploring smarter suture strategies and planning, with the aim of realizing a higher level of fully automated suture surgery.

APPENDIX A
PROOF OF THE RESULT OF (12)

In this appendix, we provide the derivation of (12) using the Lagrange multiplier method. The constrained minimization problem can be formulated as follows:

$$\begin{aligned} f(\mathbf{n}) &= \sum_{i\in\mathcal{N}(j)} ||\mathbf{q}_{i,j,k}^T\mathbf{n}||_2 = \mathbf{n}^T\left(\mathbf{Q}\mathbf{Q}^T\right)\mathbf{n} \\ \text{s.t. } & \mathbf{n}^T\mathbf{n} = \mathbf{1} \end{aligned} \tag{22}$$

The Lagrangian function is constructed as follows:

$$\mathcal{L}(\mathbf{n},\lambda) = f(\mathbf{n}) - \lambda\left(\mathbf{n}^T\mathbf{n} - 1\right) \tag{23}$$

To find the solution, we set the gradient of the Lagrangian $\mathcal{L}(\mathbf{n},\lambda)$ with respect to $\mathbf{n}$ to zero. This results in a system of equations that must be solved simultaneously:

$$\begin{cases} \frac{\partial\mathcal{L}}{\partial\mathbf{n}} = \frac{\partial}{\partial\mathbf{n}}f(\mathbf{n}) - \lambda\frac{\partial}{\partial\mathbf{n}}\left(\mathbf{n}^T\mathbf{n} - 1\right) \\ \frac{\partial\mathcal{L}}{\partial\lambda} = \mathbf{n}^T\mathbf{n} - 1 \end{cases} \tag{24}$$

For the term $\frac{\partial\mathcal{L}}{\partial\mathbf{n}}$, the expression can be simplified as:

$$\begin{aligned} & \frac{\partial}{\partial\mathbf{n}}f(\mathbf{n}) - \lambda\frac{\partial}{\partial\mathbf{n}}\left(\mathbf{n}^T\mathbf{n} - 1\right) \\ =& \left(\mathbf{Q}\mathbf{Q}^T + \mathbf{Q}^T\mathbf{Q}\right)\mathbf{n} - \lambda\left(\mathbf{I} + \mathbf{I}^T\right)\mathbf{n} \\ =& 2\mathbf{Q}\mathbf{Q}^T\mathbf{n} - 2\lambda\mathbf{n} \end{aligned} \tag{25}$$

Thus it can be deduced that:

$$\begin{cases} \dfrac{\partial\mathcal{L}}{\partial\mathbf{n}} = 0 \Longleftrightarrow \mathbf{Q}\mathbf{Q}^T\mathbf{n} = \lambda\mathbf{n} \\ \dfrac{\partial\mathcal{L}}{\partial\lambda} = 0 \Longleftrightarrow \mathbf{n}^T\mathbf{n} = 1 \end{cases} \tag{26}$$

Therefore, it can be proved that $\mathbf{n}$ is the normalized eigenvector of the smallest eigenvalue $\lambda_{\min}$ of $\mathbf{Q}\mathbf{Q}^T$.

REFERENCES

[1] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastr, "Autonomy in surgical robotics," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 4, pp. 651–679, May 2021.

[2] P. E. Dupont, B. J. Nelson, M. Goldfarb, B. Hannaford, A. Menciassi, M. K. O'Malley, N. Simaan, P. Valdastri, and G.-Z. Yang, "A decade retrospective of medical robotics research from 2010 to 2020," *Sci. Robot.*, vol. 6, no. 60, pp. 1–15, Nov. 2021.

[3] A. Sorriento, M. B. Porfido, S. Mazzoleni, G. Calvosa, M. Tenucci, G. Ciuti, and P. Dario, "Optical and electromagnetic tracking systems for biomedical applications: A critical review on potentialities and limitations," *IEEE Rev. Biomed. Eng.*, vol. 13, pp. 212–232, Sep. 2019.

[4] Y. Meng, P. Geng, M. Luo, Y. Qin, and J. Han, "An occlusion-free intraoperative active navigation system for orthopedic surgery," *IEEE Trans. Instrum. Meas.*, vol. 73, Art. no. 4005910, Mar. 2024.

[5] R. S. Decker, A. Shademan, J. D. Opfermann, S. Leonard, P. C. W. Kim, and A. Krieger, "Biocompatible near-infrared three-dimensional tracking system," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 3, pp. 549–556, Mar. 2017.

[6] Z. Han, H. Tian, T. Vercauteren, D. Liu, C. Li, and X. Duan, "Collaborative human-robot surgery for mandibular angle split osteotomy: Optical tracking based approach," *Biomed. Signal Process. Control*, vol. 93, Art. no. 106173, Mar. 2024.

[7] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein, "Electromagnetic tracking in medicine–A review of technology, validation, and applications," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1702–1725, Aug. 2014.

[8] R. J. Fonseca, *Oral and Maxillofacial Surgery-E-Book: 3-Volume Set*. Amsterdam, The Netherlands: Elsevier Health Sciences, 2017.

[9] D. A. Mitchell, *An Introduction to Oral and Maxillofacial Surgery*. Boca Raton, FL, USA: CRC Press, 2014.

[10] X. Chen, Y. Lin, C. Wang, G. Shen, S. Zhang, and X. Wang, "A surgical navigation system for oral and maxillofacial surgery and its application in the treatment of old zygomatic fractures," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 7, no. 1, pp. 42–50, Mar. 2011.

[11] X. Chen, L. Xu, Y. Sun, and C. Politis, "A review of computer-aided oral and maxillofacial surgery: Planning, simulation and navigation," *Expert Rev. Med. Devices*, vol. 13, no. 11, pp. 1043–1051, Oct. 2016.

[12] M. Zhu, B. He, J. Yu, F. Yuan, and J. Liu, "HydraMarker: Efficient, flexible, and multifold marker field generation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5849–5861, May 2023.

[13] M. S. Block and R. W. Emery, "Static or dynamic navigation for implant placement–Choosing the method of guidance," *J. Oral Maxillofacial Surg.*, vol. 74, no. 2, pp. 269–277, Feb. 2016.

[14] S. Yang, J. Chen, A. Li, P. Li, and S. Xu, "Autonomous robotic surgery for immediately loaded implant-supported maxillary full-arch prosthesis: A case report," *J. Clin. Med.*, vol. 11, no. 21, pp. 6594–6604, Nov. 2022.

[15] Y. Hu, M. Zhu, S. Wang, D. Li, F. Yuan, and J. Yu, "A novel lightweight navigation system for oral and maxillofacial surgery using an external curved self-identifying checkerboard," *IEEE Trans. Automat. Sci. Eng.*, vol. 21, no. 2, pp. 1434–1444, Apr. 2024.

[16] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao, "Augmented reality navigation with automatic marker-free image registration using 3-D image overlay for dental surgery," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1295–1304, Apr. 2014.

[17] J. Wang, H. Suenaga, L. Yang, E. Kobayashi, and I. Sakuma, "Video see-through augmented reality for oral and maxillofacial surgery," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 13, no. 2, pp. 1754–1767, Jun. 2017.

[18] 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 Trans. Mechatronics*, vol. 25, no. 2, pp. 1084–1094, Apr. 2020.

[19] D. Li, M. Zhu, S. Wang, Y. Hu, F. Yuan, and J. Yu, "A vision-based navigation system with markerless image registration and position-sensing localization for oral and maxillofacial surgery," *IEEE Trans. Instrum. Meas.*, vol. 72, Art. no. 5005811, Jan. 2023.

[20] Z. Baili, I. Tazi, and Y. Salih Alj, "StapBot: An autonomous surgical suturing robot using staples," in *Proc. Int. Conf. Multimedia Comput. Syst.*, Marrakech, Morocco, Apr. 2014, pp. 485–489.

[21] J. Wang, C. Yue, G. Wang, Y. Gong, H. Li, W. Yao, S. Kuang, W. Liu, J. Wang, and B. Su, "Task autonomous medical robot for both incision stapling and staples removal," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 3279–3285, Apr. 2022.

[22] B. Li, B. Lu, H. Lin, Y. Wang, F. Zhong, Q. Dou, and Y. Liu, "On the monocular 3D pose estimation for arbitrary shaped needle in dynamic scenes: An efficient visual learning and geometry modeling approach," *IEEE Trans. Med. Robot. Bionics*, vol. 6, no. 2, pp. 460–474, May 2024.

[23] R. Bendikas, V. Modugno, D. Kanoulas, F. Vasconcelos, and D. Stoyanov, "Learning needle pick-and-place without expert demonstrations," *IEEE Robot. Automat. Lett.*, vol. 8, no. 6, pp. 3326–3333, Jun. 2023.

[24] S. A. Pedram, P. Ferguson, J. Ma, E. Dutson, and J. Rosen, "Autonomous suturing via surgical robot: An algorithm for optimal selection of needle diameter, shape, and path," in *Proc. Int. Conf. Robot. Automat.*, Singapore, May 2017, pp. 2391–2398.

[25] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous suturing framework and quantification using a cable-driven surgical robot," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 404–417, Apr. 2021.

[26] J. Zhang, W. Wang, Y. Cai, J. Li, Y. Zeng, L. Chen, F. Yuan, Z. Ji, Y. Wang, and J. Wyrwa, "A novel single-arm stapling robot for oral and maxillofacial surgery–Design and verification," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1348–1355, Apr. 2022.

[27] S. Wang, M. Zhu, Y. Hu, D. Li, F. Yuan, and J. Yu, "Accurate detection and localization of curved checkerboard-like marker based on quadratic form," *IEEE Trans. Instrum. Meas.*, vol. 71, Art. no. 5017711, Aug. 2022.

[28] S. Wang, M. Zhu, Y. Hu, D. Li, F. Yuan, and J. Yu, "CylinderTag: An accurate and flexible marker for cylinder-shape objects pose estimation based on projective invariants," *IEEE Trans. Vis. Comput. Graph.*, early access, 2024, doi: 10.1109/TVCG.2024.3350901.

[29] J. L. Schonberger and J. -M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 4104–4113.

[30] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Proc. Int. Workshop Vis. Algorithms*, Sep. 1999, pp. 298–372.

[31] T. Zhang, J. Wang, S. Song, and M. Q.-H. Meng, "Wearable surgical optical tracking system based on multi-modular sensor fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, Art. no. 5006211, Feb. 2022.

[32] J. Wang, T. Zhang, Z. Zhang, M. Q.-H. Meng, and S. Song, "Tracking-by-registration: A robust approach for optical tracking system in surgical navigation," *IEEE Trans. Instrum. Meas.*, vol. 72, Art. no. 4010910, Sep. 2023.

Shaoan Wang received the B.E. degree in mechanical engineering from the School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree in general mechanics and foundation of mechanics with the College of Engineering, Peking University, Beijing, China. His current research interests include robot vision and visual localization.

Qiming Zhao received the B.E. degree in robot engineering from School of Mechanical Engineering & Automation, Beihang University, Beijing, China, in 2022. He is currently pursuing a master's degree in mechanical engineering with the School of Mechanical Engineering & Automation, Beihang University, Beijing, China. His current research interests include robotic motion control and force compliance control.

Dongyue Li received the B.E. degree from the School of Aerospace, Beijing Institute of Technology, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree in general mechanics and foundation of mechanics with the College of Engineering, Peking University, Beijing, China. His current research interests include robot vision and surgical robot.

Yaoqing Hu received the B.E. degree from University of Science and Technology Beijing, China, in 2018, and the M.E. degree from University of Science and Technology Beijing, China, in 2021. He is currently pursuing the Ph.D. degree in general mechanics and foundation of mechanics with the College of Engineering, Peking University, Beijing, China. His research interests include computer vision and robotics.

Mingzhu Zhu received the B.E, M.E, and D.E. degrees in the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou, China, in 2012, 2015, and 2019, respectively. From 2019 to 2021, he was a Postdoctoral Research Fellow with BIC-ESAT, College of Engineering, Peking University, Beijing, China. In 2021, he joined Fuzhou University, as an Associate Scientist. His current research interests include computer vision and image processing.

Fusong Yuan received the B.A. degree in Dentistry from Weifang Medical University, Shandong, China, in 2008, and the M.E. degree and Ph.D. degree in Prosthodontics from Peking University, Beijing, China, in 2011 and 2014, respectively.

In 2014, he joined the Peking university school and hospital of Stomatology, as a dentist, teacher, and researcher. In 2021, he was promoted to associate chief physician. His current research interest includes the application research of robotic and femtosecond laser technology in stomatology.

Jinyan Shao received the B.E. degree in measurement and control from Hunan University, Changsha, China, in 2002, and the Ph.D. degree in dynamics and control from Peking University, Beijing, China, in 2007.

After graduation, she joined IBM ResearchChina as a Research Staff Member. She focused on researching and developing innovative cross-industry solutions and intelligent systems based on big data analytics and AI technologies. In 2021, she joined College of Engineering, Peking University, as a Research Scientist. She has more than ten patents granted and has authored or coauthored more than 20 publications in referred conferences and journals. Her current research interests include multirobot system control, biomimetic robotics, underwater robots, and complex system modeling.

Junzhi Yu (Fellow, IEEE) received the B.E. degree in safety engineering and the M.E. degree in precision instruments and mechanology from the North University of China, Taiyuan, China, in 1998 and 2001, respectively, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2003.

From 2004 to 2006, he was a Postdoctoral Research Fellow with the Center for Systems and Control, Peking University, Beijing. In 2006, he was an Associate Professor with the Institute of Automation, Chinese Academy of Sciences, where he became a Full Professor in 2012. In 2018, he joined the College of Engineering, Peking University, as a Tenured Full Professor. His current research interests include intelligent robots, motion control, and intelligent mechatronic systems.