

ArthroNavi framework: stereo endoscope-guided instrument localization for arthroscopic minimally invasive surgeries

Zhongjie Long^{①, a,*} Yongting Chi^{①, a} Xiaotong Yu^{①, b} Zhouxiang Jiang^{①, a}
and Dejin Yang^{①, c,*}

^aBeijing Information Science & Technology University, School of Electromechanical Engineering, Beijing, China

^bGuang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

^cBeijing Jishuitan Hospital, Capital Medical School, 4th Clinical College of Peking University, Department of Orthopedics, Beijing, China

ABSTRACT. **Significance:** As an example of a minimally invasive arthroscopic surgical procedure, arthroscopic osteochondral autograft transplantation (OAT) is a common option for repairing focal cartilage defects in the knee joints. Arthroscopic OAT offers considerable benefits to patients, such as less post-operative pain and shorter hospital stays. However, performing OAT arthroscopically is an extremely demanding task because the osteochondral graft harvester must remain perpendicular to the cartilage surface to avoid differences in angulation.

Aim: We present a practical ArthroNavi framework for instrument pose localization by combining a self-developed stereo endoscopy with electromagnetic computation, which equips surgeons with surgical navigation assistance that eases the operational constraints of arthroscopic OAT surgery.

Approach: A prototype of a stereo endoscope specifically fit for a texture-less scene is introduced extensively. Then, the proposed framework employs the semi-global matching algorithm integrating the matching cubes method for real-time processing of the 3D point cloud. To address issues regarding initialization and occlusion, a displaying method based on patient tracking coordinates is proposed for intra-operative robust navigation. A geometrical constraint method that utilizes the 3D point cloud is used to compute a pose for the instrument. Finally, a hemisphere tabulation method is presented for pose accuracy evaluation.

Results: Experimental results show that our endoscope achieves 3D shape measurement with an accuracy of $<730 \mu\text{m}$. The mean error of pose localization is 15.4 deg (range of 10.3 deg to 21.3 deg; standard deviation of 3.08 deg) in our ArthroNavi method, which is within the same order of magnitude as that achieved by experienced surgeons using a freehand technique.

Conclusions: The effectiveness of the proposed ArthroNavi has been validated on a phantom femur. The potential contribution of this framework may provide a new computer-aided option for arthroscopic OAT surgery.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.28.10.106002](https://doi.org/10.1117/1.JBO.28.10.106002)]

Keywords: interventional systems; pose estimation; instrument localization; stereo endoscope; electromagnetic tracking; osteochondral autograft transplantation (mosaicplasty).

Paper 230162GRR received Jun. 5, 2023; revised Sep. 24, 2023; accepted Sep. 29, 2023; published Oct. 14, 2023.

*Address all correspondence to Zhongjie Long, zhongjielong@bistu.edu.cn; Dejin Yang, fafnirdj@163.com

1 Introduction

1.1 Background and Motivation

This study is motivated by the clinical need for surgical navigation in minimally invasive arthroscopic surgery. An example of minimally invasive arthroscopic surgery is osteochondral autograft transplantation (OAT),¹ which is an option for repairing focal cartilage defects in the knee joints. OAT is a useful treatment for small osteochondritis dissecans ($<2 \text{ cm}^2$)^{2,3} and works by replacing a focal cartilage defect area with one or more osteochondral autografts generally harvested from the non-weight-bearing area of the patient's healthy cartilage and bone. OAT makes small incisions and thus is usually performed arthroscopically and is minimally invasive. In contrast, mosaicplasty is a surgical technique that involves inserting three or more small plugs of healthy cartilage and bone from a non-weight-bearing site to fill larger areas of cartilage defects, forming a mosaic appearance.^{4,5} Mosaicplasty is usually performed through open incisions, which is more invasive and requires a longer recovery time. OAT can be conducted by open or arthroscopic procedures. Although an open procedure has better visibility of the surgical field and it enables surgeons to get direct access to almost all articular lesions, open OAT has not shown superior clinical outcomes. In a cadaveric study,⁶ open and arthroscopic procedures for plug placement were comparatively conducted, and results showed no significant difference regarding accuracy and precision between the two techniques. Hence, arthroscopic OAT is the most commonly employed method.

In contrast to open surgery, arthroscopic OAT uses an endoscope system with a camera, a light source, and a surgical instrument that passes through a small puncture incision on the joint of the patient undergoing arthroscopic surgery. Thus, it offers considerable benefits to patients, including less post-operative pain, reduced soft tissue damage, less blood loss, shorter recovery time, and hospital stays. However, arthroscopic OAT imposes many challenges on a surgeon's dexterity because of the well-known optical restrictions associated with a small field of view, the 30 deg optical angle of the arthroscope, and the lack of spatial awareness in a monocular arthroscope. Furthermore, numerous technical notes have reported that performing mosaicplasty arthroscopically to reduce invasiveness is an extremely demanding task. It is technically challenging to adjust cartilage thickness with three or more plugs in mosaicplasty.⁵ The congruency of the graft surface with the surrounding tissue is quite critical. If the graft surface protrudes above its surrounding, it may undergo necrosis and excessive wear,⁷ which mainly depends on the angle and depth of the graft insertion.⁸ Hence, the tubular harvesting chisel must be perpendicular to the articular cartilage surface for graft harvest and insertion to avoid differences in angulation and alignment,^{9–11} as shown in Fig. 1.

Furthermore, confusion such as hand-eye misalignment and instrument disorientation under arthroscopic guidance often occurs in surgery.¹³ Stereo vision technology can help overcome partial limitations by expanding the surgical scene with wide-angle cameras or displaying the 3D structure of the tissue. Consequently, equipping surgeons with surgical navigation assistance that eases the operational constraints of arthroscopic OAT surgery is vital. In general, position information and orientation awareness are important prerequisites for most surgical localization applications. To address this need, an ArthroNavi framework that combines electromagnetic (EM) sensing and stereo endoscopy for intraoperative instrument tracking/localization is proposed in this work.

1.2 Limitations of Prior Research

Many studies have made developmental efforts on instrument tracking/localization in minimally invasive procedures. Currently, the existing tracking methods mainly consist of three types: optoelectronic-based tracking, EM-based tracking, and image-based tracking. However, adopting the existing techniques for arthroscopic OAT and instrument localization remains challenging. In this section, we will discuss these popular methods and their limitations for arthroscopic OAT.

A common method for surgical localization is to use opto-electronic-based tracking systems.^{14,15} These navigation systems are currently imageless in orthopedic surgery and are

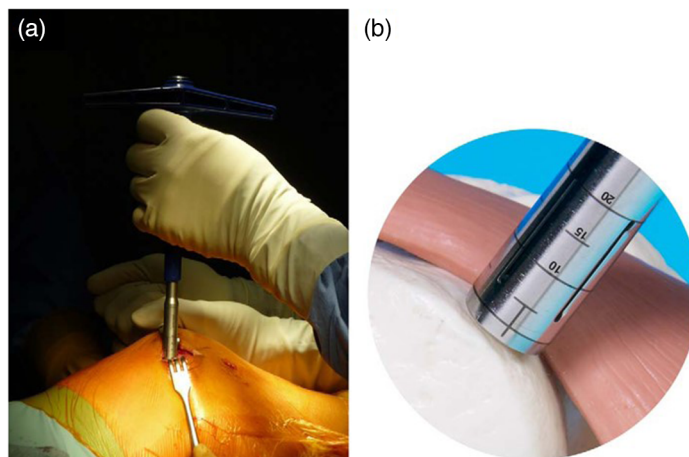


Fig. 1 (a) Harvesting of three grafts by lateral arthrotomy from the trochlea. (b) The tubular harvesting chisel should respect the dual perpendicularity of the trochlea and remain perpendicular to the cartilage surface.¹²

generally composed of a tracker, a detector, and a computer. The pose of the instrument is calculated by the tracker and a “hand–eye” calibration matrix^{16,17} that indicates the relative pose relationship between the tracker and the instrument. However, these components may seem cumbersome to inexperienced surgeons. Therefore, one or more specialized technical personnel are required to support its operation. Under this circumstance, their operations have certain constraints. For example, line-of-sight between the optical tracker and markers/detector must be cautiously maintained by the surgical team to avoid optical occlusion.¹⁸ Nevertheless, this method is mature and is widely used in clinical procedures of orthopedics.

In contrast, an EM tracking system^{19,20} generally consists of an EM field generator placed within the surgical field, tracking sensors mounted with the surgical instruments and a monitor system. The EM field generator detects signals from the tracking sensors as the surgical instruments navigate within the surgical field and uses this signal to calculate the position and orientation of the instruments in real time. The tracking sensor is capable of tracking the position and movement of the surgical instruments without occlusion, and hence this tracking approach is free of the constraints of the line-of-sight as the EM waves can penetrate through soft tissues and obstacles within the body. However, the EM-based tracking system is prone to magnetic interference, which can influence the accuracy of the tracking data.²¹ The EM trackers rigidly mounted on the instruments must be kept away from the metallic field to make the system work without any distortion.

Compared with the aforementioned two methods, image-based tracking^{22–25} for surgical localization has been reported extensively and seems to be a low-cost and easy technique in terms of system complexity. For example, the monocular structure-from-motion or simultaneous localization and mapping (SLAM) techniques^{26–28} can be integrated with current surgical setups. These methods can estimate the 3D structure from a moving monocular endoscope and simultaneously track the pose of the endoscope. The biggest advantage of SLAM-based approaches is that no optical or EM trackers are involved. Indeed, these techniques rely on the fact that the endoscopic camera provides either sparse or dense 3D reconstructions of neighbouring tissues,²⁹ which are unsuitable for arthroscopic OAT because of the following: (1) the articular surface is texture-less. An inadequate acquirement of feature points leads to a sparse point cloud that affects the accuracy of pose estimation. (2) Only the pose tracking of the endoscope is not enough for surgeons to perform OAT procedures. Registration between the endoscope, femur bone, and bone harvester must be kept consistent. (3) Another huge challenge that is unique to arthroscopic OAT is the constant extraction and insertion of the endoscope at a fast speed because of the use of various instruments, resulting in the loss of image sequences. Thus, scene initialization and mapping are required to recover the tracking during endoscope re-insertion.

1.3 Approach

An ArthroNavi framework that combines EM sensing and stereo endoscopy imaging for continuous and reliable instrument localization is presented in this study. Compared with the existing tracking method, the main merits of this framework are the continuous image scenes and no occlusion. Further, unlike our previous work,³⁰ this framework is capable of relocation tracking navigation. This development is seen as a new technical attempt for the application of navigation in specific arthroscopic OAT surgery to improve clinical outcomes. The main contributions of this study are as follows:

1. We developed a prototype of a custom-made binocular endoscope for measuring the 3D shape of the articular surface.
2. To maximize feature matching in texture-less surfaces, a speckle module that utilizes flexi-fiber for illumination in confined spaces was developed and presented in detail. The small size enables it to be embedded in the endoscope tube.
3. A robust femur coordinate was extensively introduced for visualization and localization, which was capable of dealing with intraoperative situations such as knee joint relocation or movements.

The structure of this paper is as follows. Sec. 2 introduces a prototype of a stereo endoscope, its assistance components, and the principle of imaging and locating methods. Sec. 3 shows the experimental results associated with 3D measurement and poses localization precision. Sec. 4 presents a discussion about the current study, followed by a brief conclusion of the key significance of this study.

2 Methods

2.1 Overview

In this section, the details of our proposed method are introduced thematically in terms of the workflow order. Figure 2 shows an overview of our stereo endoscope-guided intraoperative navigation framework. The proposed framework integrates EM sensing-based navigation and endoscopic vision-based navigation. It is designed with the intention of its incorporation into orthopedic workflows for operation assistance because it is a stable tracking approach that relies on a 3D points cloud obtained from a freehand stereo endoscope. Intraoperative localization of instruments can be tracked and estimated in real-time even in poor conditions that yield sparse cloud points.

2.2 Development of the Stereo Endoscope

Measuring a 3D shape of a knee joint in a confined space is challenging. Given that the imaging conditions and scene texture associated with arthroscopic imaging are poor, feature-matching-based passive vision (e.g., binocular vision) is not ideal. As a result, low matching accuracy makes sparse point cloud data, which is insufficient for pose estimation. Typically, active vision such as structured lighting in studies³⁰⁻³² can overcome texture-less problems and obtain a higher matching accuracy, but some of the studies have to insert extra probes for light illumination or move the medical endoscope to achieve a 3D shape of the test object, which seems to be inappropriate for arthroscopic OAT. Consequently, we consider that passive and active methods can be combined to develop a new endoscope, which has a high measurement accuracy and does not require relative motions between the endoscope and the test object. To solve this problem, we propose applying speckle illumination in the stereo endoscope by projecting random dots onto the measured scenes to increase the accuracy of feature matching. Moreover, the illumination probe is fixed inside the endoscopic tube so the overall size of the endoscope is sufficiently small to allow it to be used in narrow spaces.

First, we adopt two customized cameras with each packaged diameter of 3.4 mm to construct a stereo endoscope based on the binocular optical model. The two cameras are mounted side by side with parallel optical axis. Taking account of the brightness inside the knee joint, the endoscope is capable of adjusting illumination by fixing four white micro-lighting emitting diodes that sit radially around the frontend of each camera. Table 1 lists the specifications of the camera used.

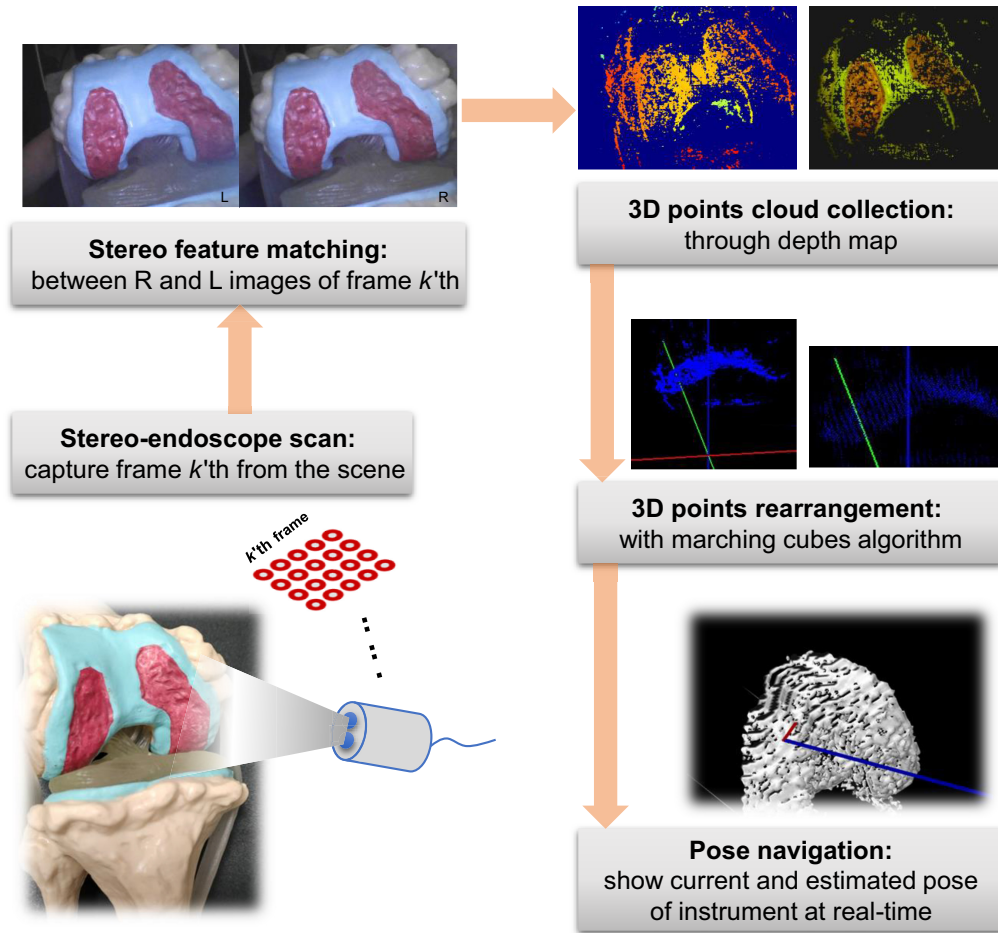


Fig. 2 Workflow of the stereo endoscope-guided navigation framework.

Table 1 Specifications of the customized camera.

Parameters	Values
Sensor	CMOS OV9734
Pixel size	1.4 μm \times 1.4 μm
Resolution	1 million pixels/720 P
Frame rate	30 frames per second
Field of view	120 deg
Depth of field	10 to 100 mm
Scan mode	Progressive

Next, a custom-designed speckle illuminator is developed and presented in detail. Figure 3(a) shows the structure diagram of the custom-designed speckle illuminator. The speckle illuminator mainly consists of three components: coupling modules, collimation lens, and a diffractive optical element (DOE). The laser ray coming from the diode projector is coupled into the proximal end surface of the imaging fiber (0.22 NA, 3.5 μm core, 900 μm cladding), wherein the coupling modules consist of a TO56 and coupling lens, as shown in Fig. 3(b). On the other side of the imaging fiber, the collimation lens (F2) is used to collimate the emergent ray prior to the DOE lens. Moreover, a diffraction process used to make the ray output can be finely tuned to produce a

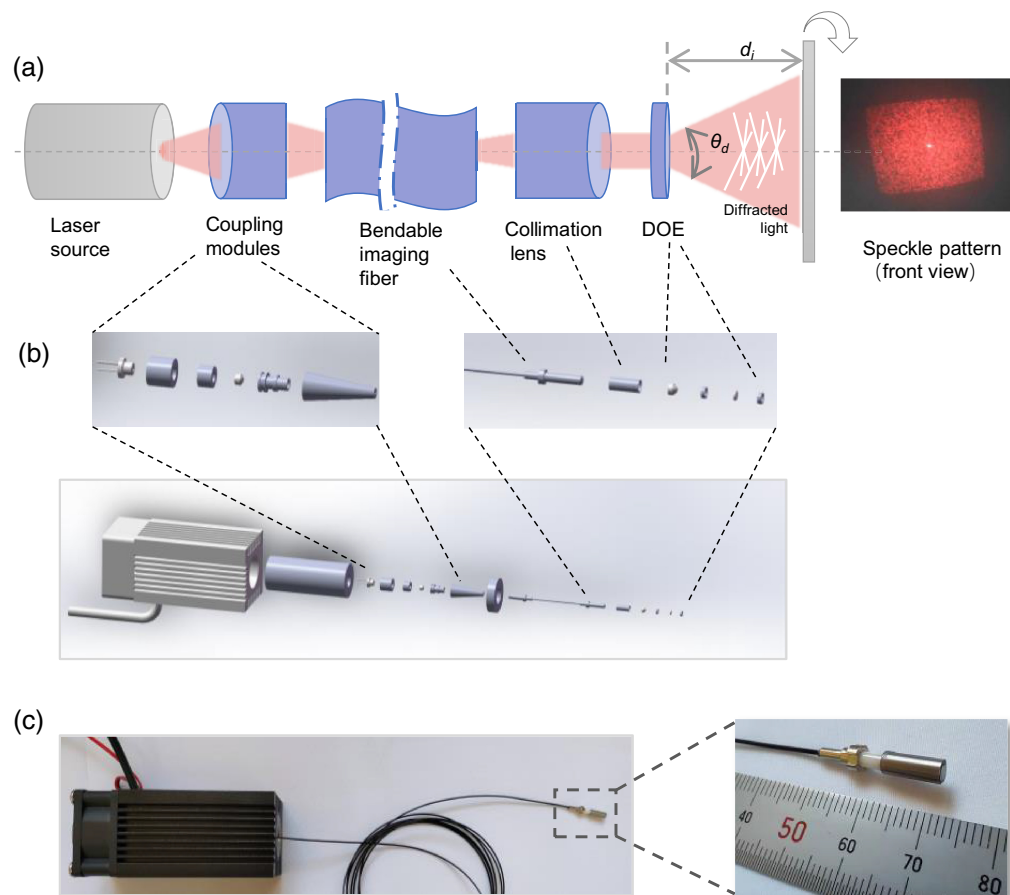


Fig. 3 Custom-designed speckle illuminator. (a) Optical design layout of the illuminator. (b) Explosion diagram of structure composition of the illuminator. (c) Photograph of the prototype and its distal end.

particular speckle pattern. Figure 3(b) shows an exploded view of the speckle illuminator, and Fig. 3(c) shows the prototype of the speckle illuminator. In this study, the illuminator is originally designed to a working distance of ~ 20 mm to meet the needs of arthroscopic OAT application. Based on our pre-experiments, the number of a random dot is set to ~ 5000 points to generate a distinct speckle pattern for camera detection. Additional parameters about the illuminator are given in Table 2.

As shown in Fig. 2, the first step in the workflow is to scan the scene and capture one proper frame k 'th by the stereo endoscope. The right and left images in frame k 'th must be corrected for radial and tangential distortions before feature matching. Thus, a standard camera calibration process proposed by Zhang³³ is conducted using the C++ platform to compute the intrinsic parameters of the two cameras. Our experiments are currently conducted without water; however, the real arthroscopic OAT surgery is performed in an aqueous environment. In that case, underwater calibration is also required as differences in the optical properties of the medium give rise to different intrinsic parameters.

For clarity, Fig. 4(a) shows the front view of the endoscope tip, which clearly describes the relative position of each component. Two cameras, a speckle illuminator, and an EM sensor are aligned in a circle with a diameter of 7.40 mm. The EM sensor is specified in detail in Sec. 2.4. A photograph of the endoscopic tip is shown in Fig. 4(b). Given that the sensor is a cylindrical shape, its correct position must be found before attaching it to the camera, otherwise, the relative orientation between the sensor and the camera cannot be obtained. Figure 4(c) shows the XOY plane of the sensor. Currently, the endoscope tip is mounted with light-cured resin to meet the needs of simulation experiments and to reduce the cost of using industrial-grade packaging.

Table 2 Specifications of the custom-designed speckle illuminator.

Component	Parameter	Value
Laser source	Wavelength	650 nm
	Laser power	0 to 30 mW
	Coupling efficiency	30%
	Working voltage	12 V
	Overall dimension	95 mm × 40 mm × 40 mm
Bendable imaging fiber	Type	Pure silica core single-mode fiber
	Fiber diameter	3.5 μm
	Numerical aperture (NA)	0.22
	Cladding	900 μm, plastic
	Patch cord	FC/PC
	Length	1000 mm
	DOE	Focal length
	Dimension	φ3.5 mm × 8 mm
	Number of speckle	~5000 points
	Divergence angle	65 deg

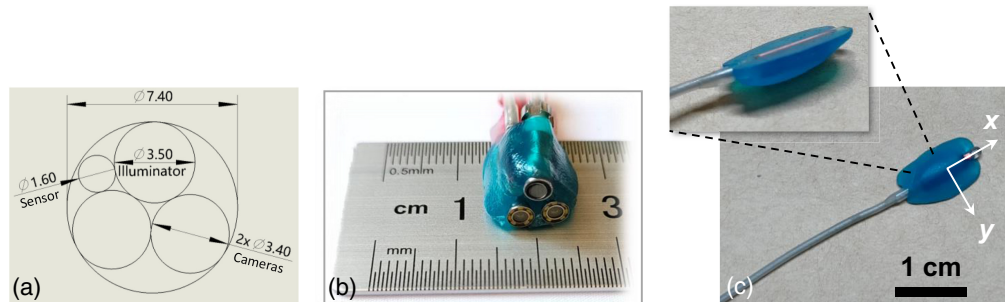


Fig. 4 Photos of the distal end of the stereo endoscope. (a) The end-face view of the endoscopic layout. (b) A close-up shot of the endoscopic tip. (c) The correct direction of the sensor's XOY plane.

2.3 Feature Matching

After scene capture, the next step is to establish a pixel-to-pixel relationship $R_{\text{match}}: l(x_p, y_p) \mapsto r(x_p, y_p)$ using epipolar lines of the two images taken by the two cameras. The semi-global matching (SGM) algorithm, which is a classical dense stereo matching approach proposed by Hirschmuller,³⁴ is adopted for this process. The SGM method is based on the idea of pixel-wise matching of mutual information and approximates a global 2D smoothness constraint by combining many 1D constraints. Although numerous improved approaches have been proposed based on the original SGM algorithm, in machine vision, SGM is preferred due to its good trade-off between precision and computation requirements. Matching of images with different exposures and lighting have been tested in the original SGM algorithm. The results indicate that the average errors for matching images with differing exposures were observed to be below 20%, and those of different illuminations were ~35%.³⁴

2.4 Patient Tracking Coordinates and Point Cloud Collection

To obtain a successional pose estimation on the knee-joint surface, the Liberty Polhemus system (LIBERTY, Polhemus, Colchester, Vermont, United States), a state-of-the-art EM-based tracking system, is used in our ArthroNavi framework. The system consists of a transmitter and up to four sensors. The transmitter produces an EM field that acts as an accurate reference for the position and orientation measurements of the sensors.

The patient tracking coordinate, as one of the highlights in this study, is proposed for the intra-operative robust 3D displaying. The difference between our approach and the previous imaging method utilized in another study³⁰ is shown in Fig. 5, which illustrates the use of different coordinate systems. Figure 5(a) shows the previous imaging method. When the femur surface is scanned and calculated by the endoscope, the surface will be incrementally displayed in the transmitter coordinate system. Let $P_i \in \mathbb{R}^3$, $i \in \{1, 2, \dots, n\}$ denote a point set in the minimally invasive surgery (MIS) scene, and ${}^c P_i$ be the corresponding point cloud produced from the endoscopic camera system. The 3D point cloud can be transformed to the transmitter's view by the following equation:

$${}^t P = {}^t_{s1} R ({}^{s1} d + {}^{s1} R \cdot {}^c P) + {}^t P_{s1}, \quad (1)$$

where ${}^t_{s1} R$ and ${}^t P_{s1}$ are the relative orientation and position of sensor-1 with respect to the transmitter coordinate system; both of them can be obtained from the Liberty EM tracking system. ${}^{s1} R$ denotes the relative orientation of the camera with respect to the sensor-1 coordinate, and ${}^{s1} d$ is the relative distance between the sensor-1 and the camera.

Although the scanned surface can be displayed based on the transmitter system, some inevitable situations may throughout the surgical operation. For example, (1) slight backward or forward movement of the femur during orthopedics operation and (2) orthopedists may knock against the transmitter. As a result, the position and orientation of the obtained 3D surface do not match that of the femur in the current position, as shown by the yellow points in Fig. 5(a). Hence, subsequent localization would not have succeeded on the femur surface.

For our proposed framework, as can be seen in Fig. 5(b), we add an additional sensor-3 to the femur. In particular, sensor-3 is mounted rigidly to minimize the influence of specimen motion. By doing so, sensor-3 can be considered as the femur coordinate that takes the role of patient tracking. Thus, in contrast with Fig. 5(a), the transmitter here will change to an intermediary, which allows us to achieve great robustness. A 3D point in the MIS scene P can be transformed into the femur coordinate as follows:

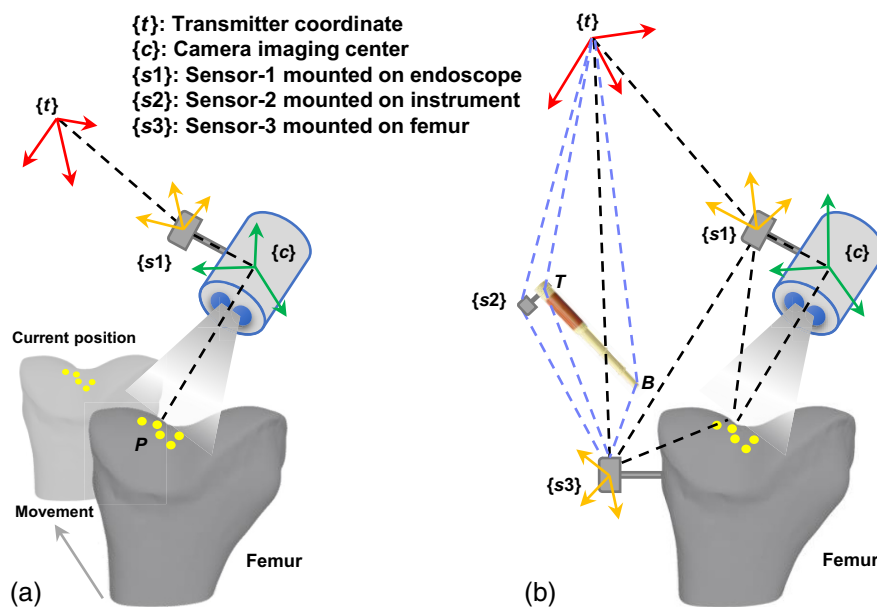


Fig. 5 Comparison of the 3D imaging coordinate. (a) The previous tracking method. (b) Our proposed ArthroNavi tracking framework.

$${}^{s3}\mathbf{P} = {}^{s3}\mathbf{R}({}_t^{s1}\mathbf{R}({}^{s1}\mathbf{d} + {}_c^{s1}\mathbf{R} \cdot {}^c\mathbf{P}) + {}_t\mathbf{P}_{s1}) + {}^{s3}\mathbf{P}_t, \quad (2)$$

where ${}^{s3}\mathbf{R}$ and ${}^{s3}\mathbf{P}_t$ are the relative orientation and position of the transmitter with respect to sensor-3. Given that ${}^{s3}\mathbf{R} = {}_t^{s3}\mathbf{R}^{-1} = {}_t^{s3}\mathbf{R}^T$ and ${}^{s3}\mathbf{P}_t = -{}^{s3}\mathbf{R}^T\mathbf{P}_{s3} = -{}_t^{s3}\mathbf{R}^T\mathbf{P}_{s3}$ are known, ${}^{s3}\mathbf{R}$ and ${}^{s3}\mathbf{P}_t$ can be calculated. By utilizing the local patient tracking coordinate as an agent, we solved three vital issues for intraoperative localization in arthroscopic OAT: (1) intraoperative femur can be rotated or moved anywhere at the correct pose, which provides the orthopedist with operation conveniences, (2) any collision with the transmitter does not affect the localization results, and (3) no pre-operative registration or initialization is required, which saves time.

2.5 3D Points Rearrangement

We rearrange the point cloud to reconstruct surfaces locally using the refined marching cubes algorithm.³⁵ Unlike the classical marching cubes algorithm,³⁶ this method is a non-interpolation approach, which decreases the computational cost. Let a point set ${}^{s3}\mathbf{P}_i \in \mathbb{R}^3$, $i \in \{1, 2, \dots, n\}$ be the point cloud represented under sensor-3 coordinate. The smooth and continuous surface S can be reconstructed by a two-step procedure: (1) a series of interval-planes are defined as follows:

$$I_i = \{x, y \in \mathbb{R}^3 | Ax + By - D_i(t) = 0\}, \quad (3)$$

where $D_i(t) = kt + \min z_i$, $k \in \{0, 1, \dots, n\}$, and $\forall z_i, D_i(t) \in [\min z_i, \max z_i]$, and $t > 0$ is the interval parameter of planes. Meanwhile, t is also the edge interval in the plane. Using Eq. (3), a cubic grid is created among the point cloud. (2) For each point located in a cube, projections along three directions will be calculated. A point will be rearranged to the adjacent vertex if the projection distance is less than half of the interval on each axis. Otherwise, the point will be rearranged to the next vertex. Accordingly, based on an eight-bit indicator (which is equal to one with a point and zero without points), we can extract the local isosurface and then reconstruct the entire surface.

2.6 Pose Computation

After surface recovery, a geometrical constraint method that utilizes the surface 3D points is used to compute a pose for the intraoperative instrument. In the case of arthroscopic OAT surgery, the orientation of the instrument is more important than its position because the position information can be obtained by the tracking system. The normal vector of every position on the reconstructed surface represents the best insertion or extraction orientation of the current instrument. Based on the proposed tracking framework, shown in Fig. 5(b), the current pose of the instrument can be shown by the vector \vec{TB} using points T and B , which are expressed as follows:

$${}^{s3}\mathbf{P}_T = {}^{s3}\mathbf{R}({}_t^{s2}\mathbf{R} \cdot {}^{s2}\mathbf{P}_T + {}_t\mathbf{P}_{s2}) + {}^{s3}\mathbf{P}_t, \quad (4)$$

and

$${}^{s3}\mathbf{P}_B = {}^{s3}\mathbf{R}({}_t^{s2}\mathbf{R} \cdot ({}^{s2}\mathbf{P}_T + \vec{TB}) + {}_t\mathbf{P}_{s2}) + {}^{s3}\mathbf{P}_t, \quad (5)$$

where ${}^{s2}\mathbf{P}_T$ and \vec{TB} can be measured beforehand.

As introduced in Sec. 2.4, 3D points P_i in the MIS scene are matched and transformed to the femur coordinate sensor-3, which is discrete character data. Thus, we approach the normal vector problem through a cross-product that computes two arbitrary vectors, \mathbf{i} and \mathbf{j} , built by a triangle. Consequently, the problem becomes that of finding an inscribed triangle Δs whose cross product is the best approximation of the normal vector \mathbf{n}_e on current position

$$\min_{\Delta s} \|\mathbf{k} - \mathbf{n}_e\|, \quad (6)$$

where $\mathbf{k} = \mathbf{ij}$. Notably, the robustness of the triangle finding is increased with a large area. However, when the triangle area $\Delta s \rightarrow 0$, the confidence level of approximation will be higher. Therefore, the parameter of the searching area in the point cloud is defined as ~ 0.6 mm in this study. Based on this geometrical constraint, the pose and normal vector can be simultaneously obtained and shown according to different positions of the 3D surface. However, even if these two vectors are overlapped, the discrepancy between measurements and theory remains because

the real normal vector is unknown. Therefore, the procedure of pose evaluation is required, as shown in the following section.

2.7 Pose Assessment Method

To evaluate the accuracy of pose estimation, we proposed an ingenious hemisphere tabulation method for pose measurement. Figure 6 shows the principle of the evaluation method. A standard hemispherical shell with a radius of 98.0 mm and a recording paper of electrocardiogram (ECG) is used for the pose test and calculation, as shown in Fig. 6(a). Since the femur profile is a free-form surface and autografts are generally harvested from a smooth area,^{4,5} a similar size of hemispherical shell is selected to match the real-life femur dimensions. In particular, a highlight of this method is that the hemispherical shell is designed to be transparent, which makes the instrument pose recorded through an optic projection technique. To this end, a specific component that imitates a bone harvester in the OAT surgery is designed using an acrylic board. The 3D structural drawing is shown in Fig. 6(b). A point laser module (dimension: 3.8 mm × 13.8 mm, 1 mW, 650 nm) is embedded at the top end of the component. The axis of laser light must be arranged co-axially with that of the component. Besides, an EM sensor-2 assisted in pose navigation is mounted close to the top end of the component, as shown in Fig. 6(c). Figure 6(d) gives a close-up view of the texture sticker. Prior to frame capture, this sticker is affixed to the hemispherical shell using the electrostatic adsorption to capture as much of feature matching as possible.

Figure 6(e) shows the definition of the spherical coordinate system. The XOY plane is located on the ECG paper, which is fixed throughout the experiment. An arc \widehat{KW} passing the sphere center is evenly divided into six sections by five points, P_1, P_2, \dots, P_5 , which are marked on the shell in advance. φ is a horizontal rotation angle with respect to x direction and is defined as $\pi/12$ during the experiment. Increasing the number of arc with an interval of angle φ , for example \widehat{KW}_2 , the sample size becomes bigger, which brings higher evaluation efficiency. Geometrically, given that the central angle of a circle is the same degree as the opposite arc, we obtain the following

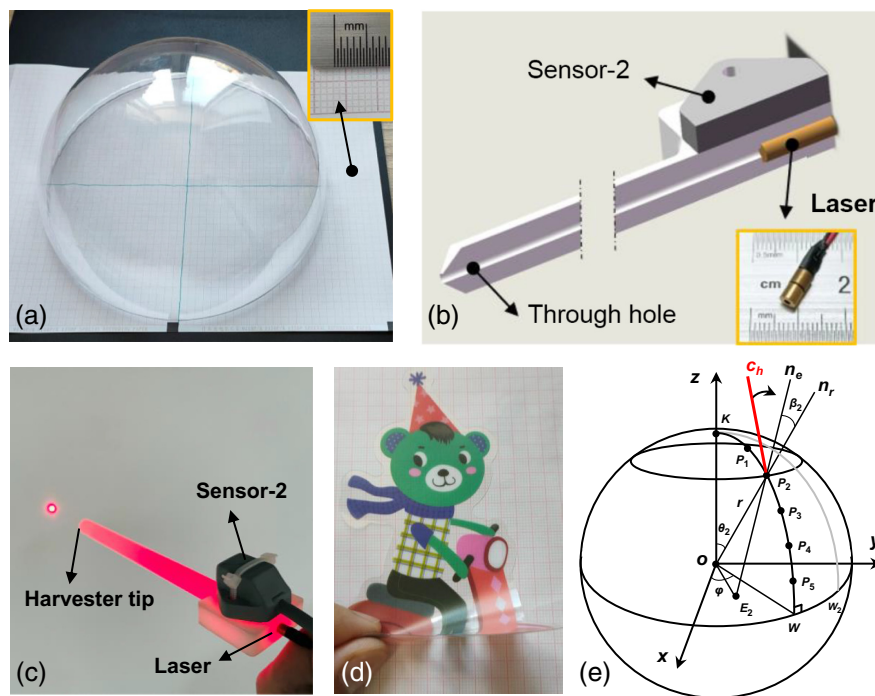


Fig. 6 Precision evaluation method for pose estimation. (a) Photograph of the hemispherical shell with an ECG paper. (b) Cutaway view of the custom-designed component harvester. (c) A laser point, passing through the hole of harvester, is projected onto a white paper. (d) A texture sticker is affixed to the hemispherical shell using static electricity. (e) The computational principle of the pose errors.

$$\theta_s = \frac{\widehat{KP}_s}{r} = \frac{s\pi}{12}, \quad (7)$$

where $s \in \{1, 2, \dots, 5\}$ is a section index on the arc and r is the spherical radius. Based on the spherical coordinate system, the 3D coordinates of the five marker point P_s on one arc can be expressed as follows:

$$\begin{bmatrix} x_{P_s} \\ y_{P_s} \\ z_{P_s} \end{bmatrix} = \begin{bmatrix} r \sin \theta_s \cos \varphi \\ r \sin \theta_s \sin \varphi \\ r \cos \theta_s \end{bmatrix}. \quad (8)$$

For each measurement of the positions on the arc, data acquisition yield a set of data samples. The pose accuracy is determined by computing these measurements.

In practice, once the frame capture is done, the texture sticker will be removed from the shell surface. When the harvester tip comes in contact with the shell surface, the navigation process is triggered, as shown in Fig. 6(e). At this stage, the current pose of the harvester c_h and the estimated normal vector n_e are displayed on the monitor during the navigation process. When these two vectors coincide, the pose represents the best fit for the current position. Meanwhile, a ray line representing the component pose is projected point to E_2 onto the ECG paper and the corresponding coordinate is recorded. Thus, based on the law of cosines, the error angle β_2 between the estimated n_e and its real n_r can be computed using the following equation:

$$\beta_2 = \arccos\left(\frac{E_2P_2^2 + r^2 - \overline{OE_2}^2}{2rE_2P_2}\right), \quad (9)$$

where \overline{VT} denotes the Euclidean distance of 3D points V and T .

3 Experiments and Results

We designed a two-part quantitative and qualitative evaluation process: (1) using a series of standard objects of flat-plane, surface, and sphere to evaluate the performance of the stereo endoscopic reconstruction error and the accuracy of the proposed ArthroNavi framework and (2) using a full-size femur model (SawBones.org) to assess the feasibility of our proposed framework.

3.1 Implementation Settings

The measurement system was implemented in a Windows 10 20H2 environment using C++ [without any graphics processing unit (GPU) acceleration] by three projects. All experiments were conducted on a laptop equipped with Intel Core 2.7 GHz CPU, 8 GB Memory, and 1 Intel HD 620 graphics card. To accelerate data reading between the three projects, a shared memory technique was adopted for inter-process communication. SGM with our proposed framework runs in real-time at 200 frames per second on average, and the 3D surface render process takes $\sim 900 \mu s$.

3.2 Precision Analysis of the Endoscope

First, to evaluate quantitatively the reconstruction accuracy of our self-development endoscope, a chessboard flat with a pattern size of 36×27 mm and the highest quality (3-start) ping-pong with a diameter of 40.09 mm were measured. The endoscope was fixed on a mount, and the measuring distance was ~ 3.7 cm for the chessboard and 4.0 cm for ping-pong. Figure 7(a) shows the corresponding objects captured by the stereo endoscope. For a clear observation, only the captured images by the left camera were shown. Figure 7(b) shows the corresponding depth maps, and Fig. 7(c) shows the 3D plot of the depth maps. Furthermore, based on the depth maps, the point cloud mapping with texture was obtained, as shown in Fig. 7(d). The depth maps obtained from the C++ project were initially in 2D format. To visualize them in 3D and enable features such as shape rotation and zooming, we utilized a third-party software. This allowed us to display the texture-mapped images in a more comprehensive manner. By doing so, we were able to select an optimal point cloud for use in the subsequent navigation process.

Besides, for comparison results, 3D imaging with a speckle pattern on the ping-pong ball was conducted, as shown in Fig. 7. The measurement parameters were the same as those without

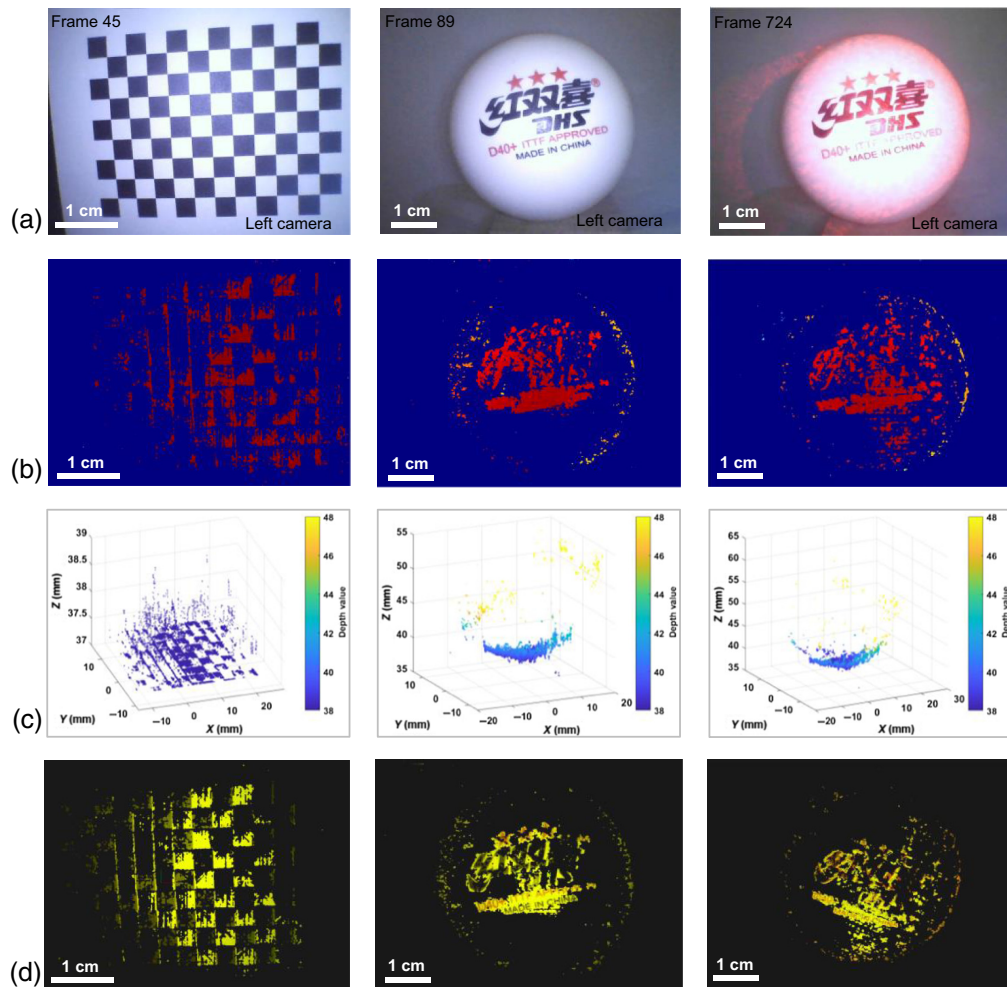


Fig. 7 3D measurement results of a chessboard flat and a ping-pong. (a) The photographs of tested objects. (b) The corresponding depth maps of the tested objects. (c) The representations of the depth maps. (d) The corresponding texture mapping of the point cloud.

speckle patterns. Figure 7(b) shows that the edge of the ball with pattern illuminated is obviously sharper, and the obtained feature points with no texture on the ball become richer. Therefore, using a speckle illuminator, the endoscope was prone to obtain a relatively big area 3D shape.

3D measurement units are commonly evaluated by a set of artifacts with a common geometry, such as planes,³⁷ spheres,³⁸ and cones.³⁹ Even a liquid crystal display was assumed as a flat plane specimen to evaluate the precision of a compact 3D measurement unit.⁴⁰ Similar to references,^{26,41,42} the method of fitting and statistic was adopted in the present study for reconstruction precision analysis. Based on the acquired 3D point cloud, the plane fitting by a polynomial was conducted to obtain the ideal plane ($R^2 = 0.9998$) as the ground truth, as shown in Fig. 8(a). The difference between the ideal plane and the measured plane was calculated to obtain the 3D measured errors. The two surfaces were aligned in the endoscope coordinate system and we get a series of (x, y) coordinate points based upon the surface area, and then compare the distance of the depth value, z , of the two surfaces. By doing so, we can obtain a point-by-point 3D measured error. Figure 8(c) shows the measurement error of the chessboard plane. Besides, the quantitative histograms of the differences were shown in Fig. 8(e). The statistical results showed that the major measurement errors were <1.3 mm with the root mean square error (RMSE) of $135.1 \mu\text{m}$. Similarly, for the 3D measurement of the ping-pong, the sphere fitting [Fig. 8(b)] was adopted to obtain the actual measurement errors, as shown in Fig. 8(d). The fitting diameter of the 3D point cloud was 39.01 mm, which was 1.08 mm deference compared with that of the ping-pong. Moreover, the RMSE of the 3D measurement accuracy was $\sim 730.8 \mu\text{m}$, as shown in Fig. 8(f).

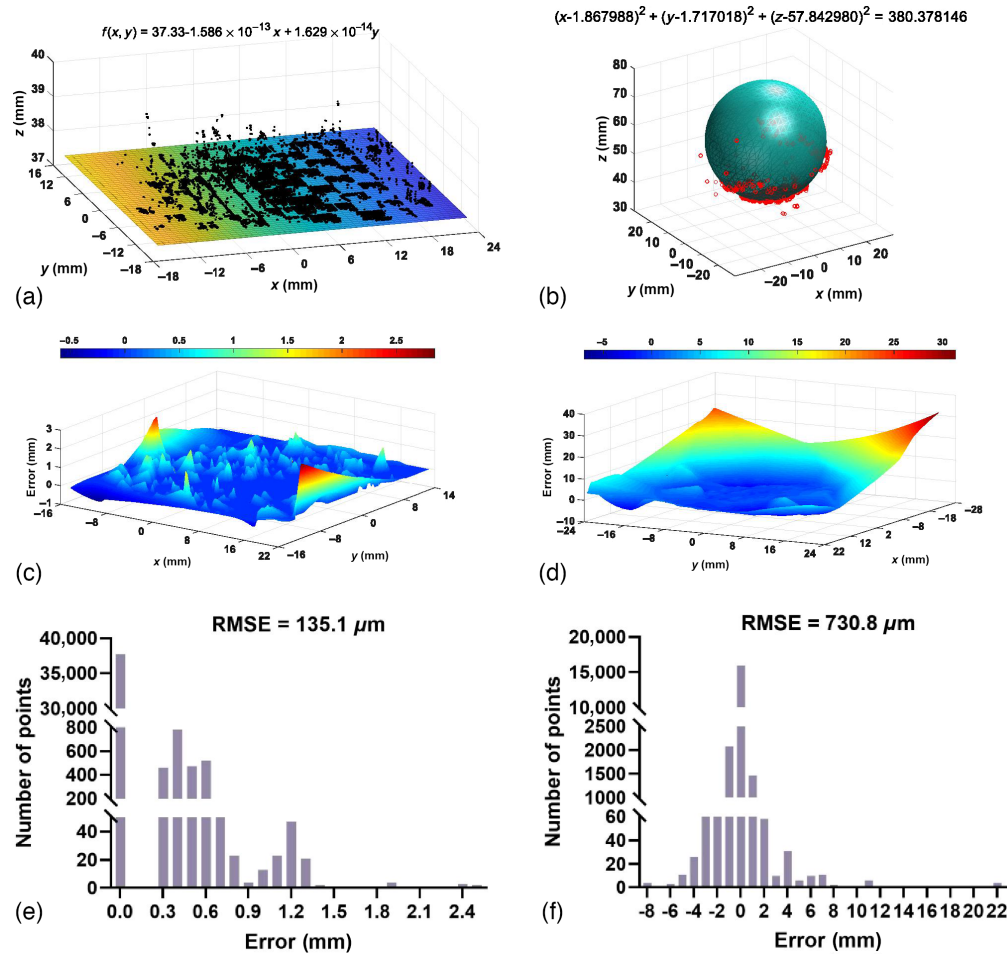


Fig. 8 Precision analysis for measuring a chessboard flat and a ping-pong. (a) The fitting results of a chessboard flat. (b) The fitting results of a ping-pong. (c), (d) The corresponding distribution of the measured errors of (a) and (b). (e), (f) The corresponding quantitative histograms of the measured errors of (a) and (b).

A comparison of reconstruction accuracy for different 3D endoscopic systems was summarized in Table 3 in ascending order of the published year. The considered aspects are the imaging technique, system setup complexity, mean error, and working distance. Note that these selected endoscopic measurement systems, the distal end, are small in size and developed for applications in MIS. Although having a small distal end, for those endoscopic measurement systems that are fixed on a desktop measuring platform⁴⁷⁻⁴⁹ are excluded for comparison. As seen in Table 3, notably, most of these endoscopic systems have a sub-millimeter accuracy in 3D reconstruction. A relative large working distance results in a smaller max error. Consequently, for handheld endoscope systems, the working distance is a main aspect that affects the imaging accuracy and decides the potential use in clinical applications. Our previous study³⁰ is more accurate than current method because it employed a monocular system structure. On the other hand, such a co-axial system is susceptible to vibration interference, which can lead to structured light (SL) projection becoming out of focus.

3.3 Precision Analysis of the Pose

After the shell surface was captured and reconstructed, pose evaluation with a freehand component was conducted. First, the endoscope was fixed at a specific position with an imaging distance of ~4.0 cm. Imaging a transparent shell with a stereo endoscope was slightly challenging. Thus, a small texture sticker with pattern was on the shell surface for scanning to maximize feature matching. Subsequently, pose navigation was carried out on the shell surface randomly.

Table 3 Precision results comparison for different 3D endoscope measurement systems.

Year	Methods	Imaging technique	System complexity	Mean error (mm)	Max error (mm)	Working distance (mm)
2006	Hayashibe et al. ³⁸	Mono + SL	Medium	0.16	1.92	150 to 160
2014	Kumar et al. ⁴³	Mono + 3D CT model	Low	1.08	1.78	Not given
2015	Edgcumbe et al. ³⁷	Mono + SL	Low	1.40	2.50	166 ± 7
2015	Yang et al. ⁴⁴	Mono + 3D US ^a image	High	0.11	0.19	~180
2015	Lin et al. ⁴⁵	Mono + SL	High	0.67	5.04	Not given
2018	Chen et al. ²⁶	Mono + SLAM	Low	2.54	Not given	Not given
2018	Lin et al. ⁴⁶	Mono + SL	High	0.64	3.19	15 to 40
2020	Sui et al. ⁴²	Stereo + SL	High	0.13	0.18	Not given
2021	Long et al. ³⁰	Mono + SL	Low	0.15	0.24	2 to 21
2023	Proposed method	Stereo + SL	Medium	0.14	4.00	37 to 40

^aUS denotes the ultrasound.

SL: Structured Light; CT: Computed Tomography.

Experimental results suggested that the mean errors of pose localization were 15.4 deg (range of 10.3 deg to 21.3 deg), with a standard deviation (SD) of 3.08 deg. Figure 9(a) shows the normal distribution of the pose estimation results. This analysis was based on 30 position estimations of 6 different static endoscope captures. For each capture, five estimations on different positions (i.e., interval angle φ) were computed. In contrast, the errors observed in a freehand technique performed by experienced surgeons during an arthroscopic surgery study⁵⁰ comparing computer-assisted navigation to the freehand technique were measured at 14.8 deg (range of 6 deg to 26 deg), with a SD of 7.53 deg. These results were derived from a hypothesis that the computer-navigated method would offer greater precision in positioning with respect to the perpendicularity of the grafts relative to the joint surface when compared to the freehand arthroscopic technique. After graft transplantation was performed using the freehand approach, positioning accuracy was assessed in a similar fashion to the navigated procedure, facilitating a direct comparison. With regard to the experimental results, our proposed method was able to achieve results within the same order of magnitude as those accomplished by experienced surgeons. Besides, instrument pose conducted by the freehand technique mainly relied on the dexterity and expertise of the surgeon, which may yield nonuniform results, whereas those in our method were automatic and identical. Nonetheless, according to the quantitative score table for guide concepts proposed by Audenaert et al.,⁵¹ the pose results obtained in our method and the freehand technique are both beyond the clinically “acceptable” range (i.e., error <4 deg and 4 mm). Therefore, achieving greater accuracy will be necessary. Figure 9(b) shows the navigation disparity between our ArthroNavi and the clinically acceptable standard.

Table 4 summarizes a comparison of pose accuracy for different instrument localization methods in descending order of the orientation accuracy. Numerous studies on pose estimation have been reported; these studies were selected for comparison because they fell in the field of bone surgery, especially in orthopedics. It should be noted that the position error in our proposed method is cited from the study of Polhemus EM tacking calibration.⁵⁷

3.4 Phantom-Based Validation

A femur model (normal size) was used for the validation of the stereo endoscope-guided navigation framework. Figure 10 shows the experimental configuration on the femur equipped with the proposed tracking method. The endoscope tip position for scene capture during the test was ~3 cm from the femur surface. Unlike the previous ping-pong ball test, the femur surface was captured with a freehand endoscope, which was in accordance with conditions used in medical applications. Please note that the hypothesis of phantom measurement was that the navigated

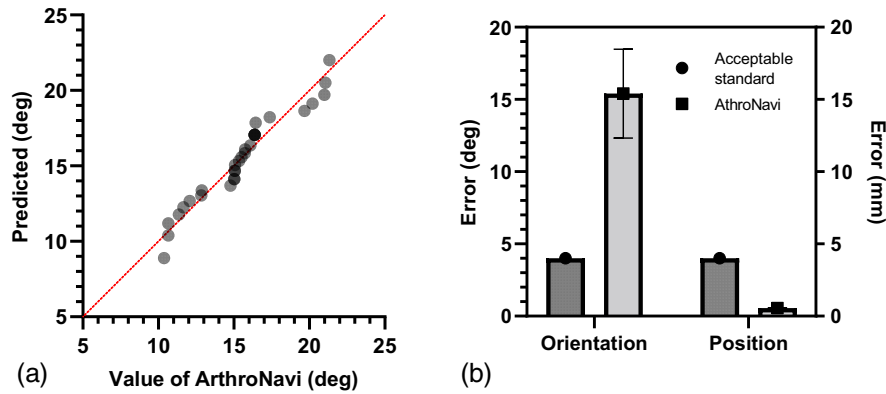


Fig. 9 Comparison results on instrument pose evaluation. (a) Normal distribution of the pose evaluation test on a hemispherical shell. (b) Navigation disparity between the ArthroNavi and the clinically acceptable standard.

Table 4 Pose accuracy comparison for different instrument localization methods. The results are given as mean \pm SD.

Year	Methods	Orientation error (deg)	Position error (mm)	Specific application
2021	Hu et al. ⁵²	1.07 \pm 0.25	4.94 \pm 0.23	Knee joint surface tracking
2018	Gadwe et al. ⁵³	1.50 \pm 0.87	1.29 \pm 0.67	Pose estimation of endoscope
2021	Hu et al. ⁵⁴	2.13 \pm 0.81	3.64 \pm 1.49	Typical knee drilling tasks
2020	Chen et al. ⁵⁵	2.55 \pm 0.49	2.54 \pm 0.15	Robot-assisted spine surgery
2020	Kügler et al. ⁵⁶	6.59 \pm 10.36	0.75 \pm 0.82	Pose estimation of a screw
2012	Benedetto et al. ⁵⁰	14.8 \pm 7.53	Not given	Grafts harvest/placement
2023	Proposed method	15.4 \pm 3.08	0.55 \pm 0.02	Grafts harvest/placement

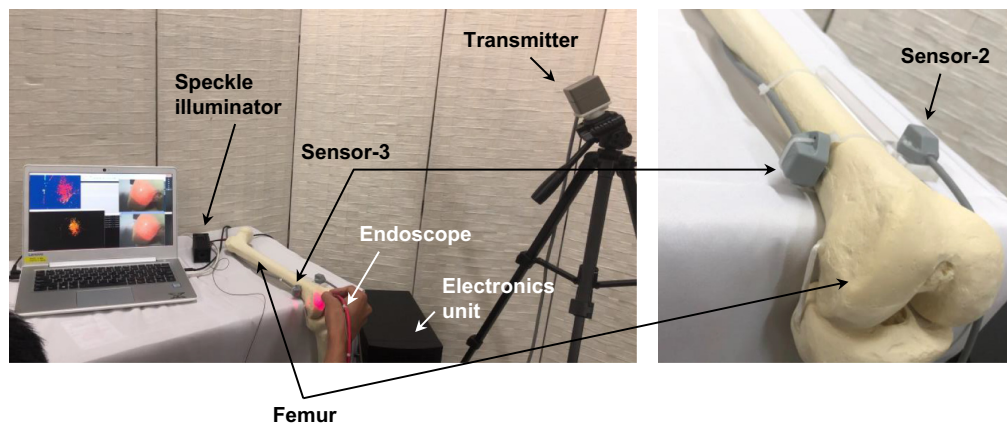


Fig. 10 Experiment setup for femur model imaging.

normal vector \mathbf{n}_e obtained from Eq. (6) would be considered the gold standard for assessing the perpendicularity of the current position of the joint surface (i.e., $\mathbf{n}_r \doteq \mathbf{n}_e$).

Based on our framework procedure, only one frame was captured for 3D imaging, as shown in Fig. 11(a). Owing to the speckle pattern illumination, the femur surface was imaged successfully in a few seconds. The corresponding depth map and a side view of the 3D point cloud were shown in Figs. 11(b) and 11(c). Once the frame was captured and saved, the endoscope task has been completed. The scene capture was a static scan without any trajectory movement. The point

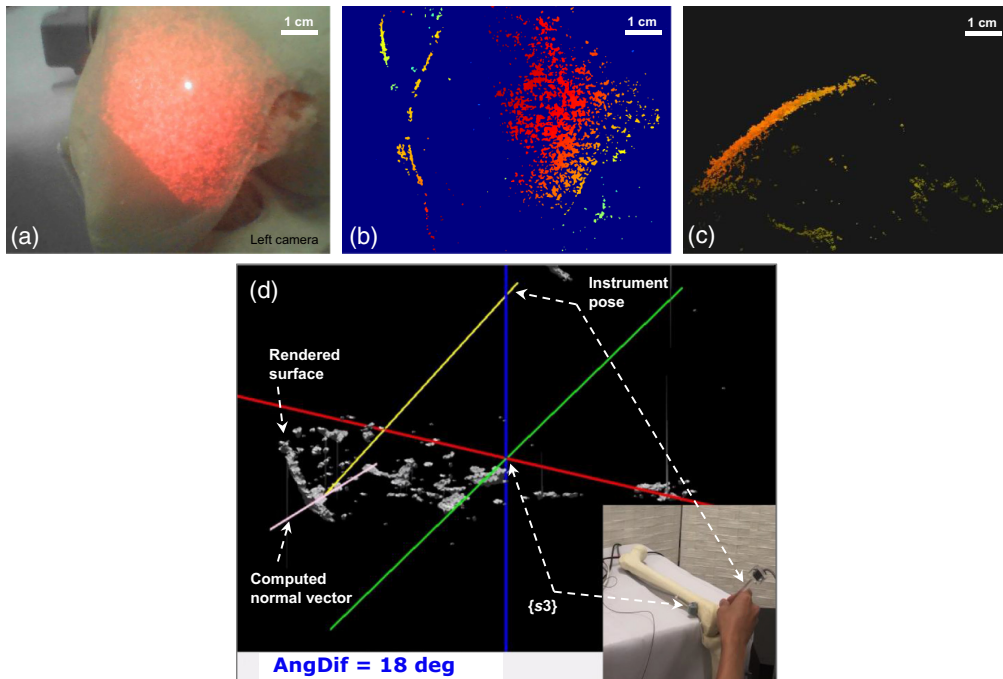


Fig. 11 Experimental results on a femur model. (a) A capture of the femur surface with a speckle pattern. (b) The corresponding depth map. (c) Side view of the 3D point cloud. (d) Snapshot from a real-time navigation video of the femur model. The instrument pose (yellow line) was synchronized with that instrument held in hand (Video 1, MP4, 7.78 MB [URL: <https://doi.org/10.1117/1.JBO.28.10.106002.s1>]).

cloud data were loaded into another C++ project that uses OpenGL for 3D surface rendering. Figure 11(d) shows a snapshot of the reconstructed surface and the instrument localization. The surface and the instrument pose were displayed based on the sensor-3 coordinate system, which was able to anti-shifting positioning. The number displayed in AngDif represents the angle between the instrument's current pose c_h and the estimated normal vector n_e used for pose adjustment. Furthermore, the graphical user interface was capable of shape rotation and zooming ability. A real-time navigation video was provided by authors in Video 1. Given the showcasing based on the femur coordinate that was introduced in Sec. 2.4, our proposed method was capable of movement tracking navigation.

4 Discussion

The current work introduces a method to create a stereo image with a self-developed endoscope and generate an image-based pose localization, with the intention for this system to be used for arthroscopic OAT surgery. The current framework can work with 30 frames per second on a laptop without a supporting GPU.

Various algorithms are available for feature matching, and we utilized the classical one proposed by Hirschmuller, which is more accurate and faster than other improved algorithms. In recent years, a number of 3D reconstruction algorithms^{58,59} based on convolutional neural network (CNN) have been presented. These methods are capable of real-time camera tracking and dense mapping after the model is trained by a huge database of labeled images. An ideal scenario is that such techniques could be integrated into the endoscope and correlate with the positions of arthroscopic surgical instruments. However, the feature in the knee joint scene is inherently poor, and model training based on a huge image database requires manual labeling of the images, which is impractical and time-consuming. Hence, the CNN-based tracking method was not our choice for image mapping.

EM-based tracking enable localization of pre-operative instruments within a patient's body without line of sight. The EM tracking accuracy and robustness is a challenge in the clinical application. However, a recent study that applied the well-established standardized assessment

protocol⁶⁰ to the Polhemus EM tracker demonstrated that the mean orientation error was found to be 0.1 deg in a laboratory environment, and the distance accuracy stayed in the sub-millimeter range on an average of 0.55 ± 0.02 mm. Precision and orientation accuracy do not seem to be affected by instrument tracking.⁵⁷ The source of error in instrument locating stemmed from the error in endoscope images. No cumulative error was yielded during the entire navigation process. Furthermore, the pose estimation method that searched three 3D points in a ~ 0.6 mm diameter circle was considered precise. This search method even works in the current sparse 3D point cloud. Thus, improving the quality of the 3D reconstructed surface would give a better localization result. To this end, we have several choices: (1) using a computational mask to filtrate image noises or abnormal points and (2) replacing the customized cameras with commercially available apparatus because customized cameras lack stability and synchronization.

While the proposed framework was developed based on the consideration of clinical requirements of arthroscopic procedures, it could be used for some general applications in diagnostic and endoscopic interventions. For example, the framework could be applied to instrument alignment and bone harvesting in femoral head replacement or entire hip replacement.

4.1 Limitations and Future Works

The main limitations of the proposed method include the following: (1) although the custom-made monocular camera has a good imaging quality at a distance of 1 cm, the optimal imaging distance of the stereo endoscope (4 cm) is beyond that of arthroscopic practice application. (2) Only a part of the femur shape was imaged and navigated for surgeons, which may lead to direction absence. (3) In the current framework, although the localization error of 15.4 deg was the same level as that of a freehand technique (14.8 deg), the localization accuracy is not so satisfied because of the sparse 3D point cloud that came from the noise in endoscope images.

Stereoscopic-image-based navigation is challenging and achievable and will be investigated in future work. One possible solution to improve imaging accuracy is to adopt higher-definition cameras and use a shared input port to solve the frame alignment problem and yield more accurate feature matching. The point number of 5000 for the current speckle module remains excessive in consideration of the ideal imaging distance of an endoscope in practice application. Based on our experience, ~ 2500 points are suitable for the ideal imaging distance of 2 to 3 cm, which enables cameras to detect an unambiguous pattern. Alternatively, the colored checkerboard pattern³⁷ or randomly distributed spots with different colors⁴⁶ are potential for better imaging.

5 Conclusion

This study contributes to existing clinical needs by developing a practical instrument localization approach that is non-disruptive to the operation process. Particularly, it proposes a complete framework for intraoperative navigation by combining reconstructed 3D surface and external trackers to bridge the gap in the application of existing tracking methods to OAT surgery. The pose localization method was validated by standard models and phantom femur. The 3D surface reconstruction is promising, and the pose navigation is operated in real-time. We hope that this prototypical framework can enlighten a new computer-aided direction for the treatment of cartilage damage in the knees.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Data, Materials, and Code Availability

All relevant data, materials, and software code used in this research are available upon request from the corresponding authors.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 52005046), the Project of Cultivation for Young Top-Notch Talents of Beijing Municipal

Institutions (Grant No. BPHR202203232), the Beijing Natural Science Foundation (Grant No. L212040), and partially by the National Natural Science Foundation of China (Grant Nos. 52175452 and 52275517). The authors wish to express their gratitude to Yinglong Li, a dentist at Beijing Chao-Yang Hospital of Capital Medical University, for his invaluable guidance in the mounting procedure of the endoscope tip using light-cured resin.

References

1. Y. Matsusue, T. Yamamuro, and H. Hama, "Arthroscopic multiple osteochondral transplantation to the chondral defect in the knee associated with anterior cruciate ligament disruption," *Arthrosc.: J. Arthrosc. Related Surg.* **9**(3), 318–321 (1993).
2. B. C. Emmerson et al., "Fresh osteochondral allografting in the treatment of osteochondritis dissecans of the femoral condyle," *Am. J. Sports Med.* **35**(6), 907–914 (2007).
3. D. Louahem et al., "Mosaicplasty for femoral osteochondritis dissecans," *Orthop. Traumatol.Surg. Res.* **102**(2), 247–250 (2016).
4. R. Chen et al., "Multiphasic scaffolds for the repair of osteochondral defects: outcomes of preclinical studies," *Bioact. Mater.* **27**, 505–545 (2023).
5. K. Kizaki et al., "Arthroscopic versus open osteochondral autograft transplantation (mosaicplasty) for cartilage damage of the knee: a systematic review," *J. Knee Surg.* **34**(1), 94–107 (2019).
6. D. Koulalis et al., "Open versus arthroscopic mosaicplasty of the knee: a cadaveric assessment of accuracy of graft placement using navigation," *Arthroscopy* **31**(9), 1772–1776 (2015).
7. A. Shekhar et al., "Mid-term outcomes of arthroscopic osteochondral autograft transplantation for focal chondral defects of the knee," *J. Arthrosc. Surg. Sports Med.* **2**(1), 41–46 (2021).
8. A. F. Mavrogenis et al., "Computer-assisted navigation in orthopedic surgery," *Orthopedics* **36**(8), 631–642 (2013).
9. M. Bordes et al., "Autologous osteochondral transplantation for focal femoral condyle defects: comparison of mosaicplasty by arthrotomy vs. arthroscopy," *Orthop. Traumatol.: Surg. Res.* **108**(3), 103102 (2022).
10. A. J. Guzman et al., "Arthroscopic osteochondral autograft transfer system procedure of the lateral femoral condyle with donor-site backfill using osteochondral allograft plug," *Arthrosc. Tech.* **10**(12), e2683–e2689 (2021).
11. D. Wan et al., "Results of the osteochondral autologous transplantation for treatment of osteochondral lesions of the talus with harvesting from the ipsilateral talar articular facets," *Int. Orthop.* **46**(7), 1547–1555 (2022).
12. H. Robert, "Chondral repair of the knee joint using mosaicplasty," *Orthop. Traumatol.: Surg. Res.* **97**(4), 418–429 (2011).
13. L. Ma et al., *3D Visualization and Augmented Reality for Orthopedics*, pp. 193–205, Springer, Singapore (2018).
14. M. J. O'Malley and B. A. Klatt, "Computer-assisted navigation-total knee arthroplasty," *Oper. Tech. Orthop.* **22**(4), 176–181 (2012).
15. C. Signorelli et al., "Validation of an optical, computer-assisted technique for intraoperative tracking of 3-dimensional Canine Stifle joint motion," *Open Vet. J.* **10**(1), 86–93 (2020).
16. S. Lee et al., "Comparative study of hand-eye calibration methods for augmented reality using an endoscope," *J. Electron. Imaging* **27**(4), 043017 (2018).
17. H. Luo et al., "Augmented reality navigation for liver resection with a stereoscopic laparoscope," *Comput. Methods Prog. Biomed.* **187**, 105099 (2020).
18. K. Cleary and T. M. Peters, "Image-guided interventions: technology review and clinical applications," *Ann. Rev. Biomed. Eng.* **12**(1), 119–142 (2010).
19. V. Lahanas, C. Loukas, and E. Georgiou, "A simple sensor calibration technique for estimating the 3D pose of endoscopic instruments," *Surg. Endosc.* **30**, 1198–1204 (2016).
20. J. Pagador et al., "Augmented reality haptic (ARH): an approach of electromagnetic tracking in minimally invasive surgery," *Int. J. Comput. Assist. Radiol. Surg.* **6**, 257–263 (2011).
21. E. B. Mark et al., "Ambulatory assessment of colonic motility using the electromagnetic capsule tracking system: effect of opioids," *Neurogastroenterol. Motil.* **32**(3), 1–10 (2020).
22. M. Allan et al., "3-D pose estimation of articulated instruments in robotic minimally invasive surgery," *IEEE Trans. Med. Imaging* **37**(5), 1204–1213 (2018).
23. J. Hein et al., "Towards markerless surgical tool and hand pose estimation," *Int. J. Comput. Assist. Radiol. Surg.* **16**, 799–808 (2021).
24. D. Wesierski and A. Jezierska, "Instrument detection and pose estimation with rigid part mixtures model in video-assisted surgeries," *Med. Image Anal.* **46**, 244–265 (2018).
25. X. Luo et al., "Monocular endoscope 6-DoF tracking with constrained evolutionary stochastic filtering," *Med. Image Anal.* **89**, 102928 (2023).
26. L. Chen et al., "SLAM-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality," *Comput. Methods Prog. Biomed.* **158**, 135–146 (2018).

27. A. Marmol, P. Corke, and T. Peynot, "ArthroSLAM: multi-sensor robust visual localization for minimally invasive orthopedic surgery," in *IEEE/RSJ Int. Conf. Intell. Rob. and Syst. (IROS)*, pp. 3882–3889 (2018).
28. N. Mahmoud et al., "Live tracking and dense reconstruction for handheld monocular endoscopy," *IEEE Trans. Med. Imaging* **38**(1), 79–89 (2019).
29. B. Lin et al., "Video-based 3D reconstruction, laparoscope localization and deformation recovery for abdominal minimally invasive surgery: a survey," *Int. J. Med. Rob. Comput. Assisted Surg.* **12**(2), 158–178 (2016).
30. Z. Long et al., "Development of an ultracompact endoscopic three-dimensional scanner with flexible imaging fiber optics," *Opt. Eng.* **60**(11), 114108 (2021).
31. M. Hayashibe, N. Suzuki, and Y. Nakamura, "Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management," *Med. Image Anal.* **10**(4), 509–519 (2006).
32. N. T. Clancy et al., "Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging," *Biomed. Opt. Express* **2**(11), 3119–3128 (2011).
33. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).
34. H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 328–341 (2008).
35. Z. Long and K. Nagamune, "A marching cubes algorithm: application for three-dimensional surface reconstruction based on endoscope and optical fiber," *Information* **18**(4), 1425–1437 (2015).
36. W. Lorensen and H. Cline, "Marching cubes: a high resolution 3D surface construction algorithm," *ACM SIGGRAPH Comput. Graphics* **21**, 163–169 (1987).
37. P. Edgumbe et al., "Pico Lantern: surface reconstruction and augmented reality in laparoscopic surgery using a pick-up laser projector," *Med. Image Anal.* **25**(1), 95–102 (2015).
38. M. Hayashibe, N. Suzuki, and Y. Nakamura, "Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management," *Med. Image Anal.* **10**(4), 509–519 (2006).
39. P. Rachakonda, B. Muralikrishnan, and D. Sawyer, "Sources of errors in structured light 3D scanners," *Proc. SPIE* **10991**, 1099106 (2019).
40. M. Fujigaki, T. Sakaguchi, and Y. Murata, "Development of a compact 3D shape measurement unit using the light-source-stepping method," *Opt. Lasers Eng.* **85**, 9–17 (2016).
41. W. Yin et al., "Single-shot 3D shape measurement using an end-to-end stereo matching network for speckle projection profilometry," *Opt. Express* **29**(9), 13388–13407 (2021).
42. C. Sui et al., "A real-time 3D laparoscopic imaging system: design, method, and validation," *IEEE Trans. Biomed. Eng.* **67**(9), 2683–2695 (2020).
43. A. Kumar et al., "Stereoscopic visualization of laparoscope image using depth information from 3D model," *Comput. Methods Prog. Biomed.* **113**(3), 862–868 (2014).
44. L. Yang et al., "Vision-based endoscope tracking for 3D ultrasound image-guided surgical navigation," *Comput. Med. Imaging Graphics* **40**, 205–216 (2015).
45. J. Lin, N. T. Clancy, and D. S. Elson, "An endoscopic structured light system using multispectral detection," *Int. J. Comput. Assist. Radiol. Sur.* **10**, 1941–1950 (2015).
46. J. Lin et al., "Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks," *Med. Image Anal.* **48**, 162–176 (2018).
47. H. M. Park and K. N. Joo, "Endoscopic precise 3D surface profiler based on continuously scanning structured illumination microscopy," *Curr. Opt. Photonics* **2**(2), 172–178 (2018).
48. Y. Escamilla and F. Otani, "Three-dimensional surface measurement based on the projected defocused pattern technique using imaging fiber optics," *Opt. Commun.* **390**, 57–60 (2017).
49. J. Schlobohm, A. Pösch, and E. Reithmeier, "A raspberry pi based portable endoscopic 3D measurement system," *Electronics* **5**(3), 43 (2016).
50. P. Di Benedetto et al., "Arthroscopic mosaicplasty for osteochondral lesions of the knee: computer-assisted navigation versus freehand technique," *Arthrosc.: J. Arthrosc. Relat. Surg.* **28**(9), 1290–1296 (2012).
51. E. Audenaert et al., "A custom-made guide for femoral component positioning in hip resurfacing arthroplasty: development and validation study," *Comput. Aided Surg.* **16**(6), 304–309 (2011).
52. X. Hu, A. Nguyen, and F. R. y Baena, "Occlusion-robust visual markerless bone tracking for computer-assisted orthopedic surgery," *IEEE Trans. Instrum. Meas.* **71**, 1–11 (2021).
53. A. Gadwe and H. Ren, "Real-time 6DOF pose estimation of endoscopic instruments using printable markers," *IEEE Sens. J.* **19**(6), 2338–2346 (2018).
54. X. Hu, H. Liu, and F. R. Y. Baena, "Markerless navigation system for orthopaedic knee surgery: a proof of concept study," *IEEE Access* **9**, 64708–64718 (2021).
55. L. Chen et al., "Research on the accuracy of three-dimensional localization and navigation in robot-assisted spine surgery," *Int. J. Med. Rob. Comput. Assist. Surg.* **16**(2), e2071 (2020).
56. D. Kügler et al., "i3PosNet: instrument pose estimation from x-ray in temporal bone surgery," *Int. J. Comput. Assist. Radiol. Surg.* **15**, 1137–1145 (2020).

57. A. M. Franz et al., “Polhemus EM tracked micro sensor for CT-guided interventions,” *Med. Phys.* **46**(1), 15–24 (2019).
58. L. Koestler et al., “TANDEM: tracking and dense mapping in real-time using deep multi-view stereo,” in *Conf. Rob. Learn.*, PMLR (2021).
59. R. Mur-Artal and J. D. Tardós, “ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras,” *IEEE Trans. Rob.* **33**(5), 1255–1262 (2017).
60. J. B. Hummel et al., “Design and application of an assessment protocol for electromagnetic tracking systems,” *Med. Phys.* **32**(7), 2371–2379 (2005).

Zhongjie Long received his BE degree in vehicle engineering from South China University of Technology in 2010. He received his ME degree in mechanical engineering in 2013 from Beijing Information Science & Technology University, and his PhD in advanced interdisciplinary science and technology in 2016 from the University of Fukui, Japan. Currently, he is an associate professor in the School of Electromechanical Engineering at Beijing Information Science & Technology University. He is interested in computer-assisted surgery systems, 3D endoscopic imaging, and tools tracking.

Yongting Chi received his BE degree in mechanical engineering from Zhuhai College of Jilin University in 2022. He is currently pursuing his ME degree at Beijing Information Science & Technology University. His research interests include endoscope development and 3D imaging.

Xiaotong Yu received her PhD in clinical medicine from Beijing University of Chinese Medicine in 2018. She is now an attending physician at Guang’anmen Hospital, China Academy of Chinese Medical Sciences.

Zhouxiang Jiang received his BS degree in mechanical engineering from Beijing Information Science & Technology University in 2008, his MS degree in mechanical engineering from China University of Geosciences in 2011, and his PhD in mechanical engineering from Huazhong University of Science and Technology in 2016. He is currently an associate professor at Beijing Information Science & Technology University. His research interests include robot calibration methods and geometric errors of five-axis machine tools.

Dejin Yang is a deputy consultant and associate professor in orthopaedics, working in the Department of Adult Joint Reconstructive Surgery in Beijing Jishuitan Hospital at the 4th Clinical College of Peking University. He practices in joint surgeries, and he is especially experienced in the treatment and research on osteonecrosis and osteoarthritis of adult hips and knees.