Three-dimensional multimodal medical imaging system based on freehand ultrasound and structured light

Jhacson Meza,[a] Sonia H. Contreras-Ortiz,[a] Lenny A. Romero,[b] and Andres G. Marrugo[a,*]

[a]Universidad Tecnológica de Bolívar, Facultad de Ingeniería, Cartagena, Colombia
[b]Universidad Tecnológica de Bolívar, Facultad de Ciencias Básicas, Cartagena, Colombia

Abstract. We propose a three-dimensional (3D) multimodal medical imaging system that combines freehand ultrasound and structured light 3D reconstruction in a single coordinate system without requiring registration. To the best of our knowledge, these techniques have not been combined as a multimodal imaging technique. The system complements the internal 3D information acquired with ultrasound with the external surface measured with the structured light technique. Moreover, the ultrasound probe's optical tracking for pose estimation was implemented based on a convolutional neural network. Experimental results show the system's high accuracy and reproducibility, as well as its potential for preoperative and intraoperative applications. The experimental multimodal error, or the distance from two surfaces obtained with different modalities, was 0.12 mm. The code is available in a Github repository. © *2021 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.OE.60.5.054106]

1 Introduction

Conventional medical procedures rely heavily on the physician's visualization skills and expertise. These can be enhanced by modern multimodal imaging technologies to perform complex tasks, such as surgical planning, navigation, and intraoperative decision making.[1] However, most intraoperative systems still rely on two-dimensional (2D) information, although the task is essentially three-dimensional (3D). Moreover, the existing multimodal 3D imaging technologies are either too expensive and bulky to be used in routine procedures[2] or not free of ionizing radiation, as in the case of computed tomography (CT).[1]

There are suitable alternatives, such as ultrasound (US) imaging, which is free of ionizing radiation, portable, and low-cost and allows for real-time visualization.[3] Furthermore, freehand scanning with a US probe provides a flexible and convenient tool for clinicians to reconstruct 3D images of the region-of-interest.[4,5] However, 3D US does not provide information about the external anatomy, such as that provided by CT or magnetic resonance imaging (MRI), which is essential for surgical planning and navigation.

Recently, digital light processing (DLP) projectors have enabled fast and accurate structured light (SL) systems for 3D surface reconstruction.[6] Although primarily used in the manufacturing industry,[7] DLP-based SL systems have found new uses in biomedical applications, such as surgical navigation,[8] real-time 3D measurement otoscopy,[9] and fluorescence imaging and topography for intraoperative imaging.[10] It is by far the most adequate technique for obtaining a digital representation of the external surface. Moreover, to the best of our knowledge, freehand US and SL have not been used simultaneously as a multimodal imaging technique. It is the programmable nature of the DLP-based SL system that opens new possibilities for such a multimodal technique with freehand US. This combination can be useful for percutaneous interventions such as biopsies, ablations, or drainages that are mainly guided using 2D imaging techniques, specifically with US, due to its versatility.[11,12] For these percutaneous procedures, it would be

---

helpful to have the 3D model of the external anatomy to improve image interpretation and spatial understanding of the US images.

In this paper, we propose a multimodal imaging system based on freehand US and SL to acquire the internal and external features of a zone of interest in the same coordinate system. The proposed system is aimed mostly for percutaneous interventions as a preoperative and intraoperative imaging tool. In this way, our system can be used for treatment planning, intervention planning with preoperative data, and procedure guidance due to its ability to easily update previously acquired data. Furthermore, it may provide a low-cost solution for surgical planning in rural or low-resource settings.[13] Our approach considers optical tracking of the US probe and the SL 3D reconstruction using a pair of cameras and a DLP projector, all calibrated to a global coordinate system. Therefore, no registration is needed for multimodal visualization, an often challenging procedure.[14] Moreover, we developed a deep learning US probe tracking algorithm for precise tracking under varying illumination conditions and motion blur. The experimental results show the potential of the proposed system as a multimodal medical imaging technique and a surgical navigation tool.

2 Related Work

US and SL are powerful, cost-effective imaging techniques that offer great flexibility and potential solutions to many medical procedures that require precise positioning and guidance. US is highly used for both open and minimally invasive procedures with an array of different probes and newly developed contrast agents.[15] Because of many of these advantages, US imaging has been combined with other techniques, such as fluorescence molecular tomography,[16] MRI for intraoperative guidance,[17,18] fluorescence lifetime imaging for oral cancer evaluation,[19] and intravascular photoacoustic imaging for characterization of atherosclerosis.[20] In summary, US is used so often with other modalities because it provides the clinicians with interactive feedback.[21]

Similarly, SL has also been combined with other techniques to take advantage of its ability to probe the surface of the region of interest. For example, it has been combined with fluorescence imaging to obtain a 3D fluorescence topography in which the DLP projector is used as an excitation source.[10] There are other instances in which SL has been used in a multimodal system, such as with micro-CT for evaluating excised breast tissue[22] and with a PET system for head motion tracking.[23] Its capacity for accurate, fast 3D surface reconstruction is unparalleled with competing techniques.

There are two cases in which US and SL systems have been used separately to address the same problem: scoliosis evaluation and breast cancer detection. In the first case, Pino-Almero et al.[24] measured the topographic changes in the back surface obtained through an SL system to monitor idiopathic scoliosis. For the same type of assessment, Cheung et al.[25] developed a 3D freehand US system to measure spine curvature based on bony landmarks of the spine extracted from a stack of US scans. In the second case, US is known to be an excellent technique for breast cancer detection.[26,27] Moreover, there is some evidence obtained with SL systems linking breast surface variation to cancer.[28] These examples give some insight into why it would be beneficial to have a multimodal freehand US and SL system.

There have been several approaches in which US and SL have been jointly used to improve or solve orientation or image guidance problems in medical imaging, although none so far as a multimodal imaging technique. For example, Horvath et al.[29] used an SL system composed of a camera and two lasers attached to a US probe to determine the orientation of the surface relative to the US transducer for solving the Doppler ambiguity problem from arteries and veins running parallel to the surface. Basafa et al.[30] used a stereo vision system mounted on the US transducer to aid a US and computer tomography multimodal system for image guidance. Similarly, there have been many attempts at mounting a camera on the US probe for sensorless freehand US.[31,32] However, the implementation is not problem-free due to insufficient tracking features in the skin or surrounding regions—not to mention the bulkier US probe resulting from additional instruments on board.

Finally, US and SL techniques have often been used to solve similar problems in medical imaging. However, considering that they provide complementary information (US internal features

and SL external surface), we find it necessary to develop a system with both techniques highly integrated as a multimodal imaging system. It is worth noting that the DLP projector in the SL system enables features such as active surgical guidance and telementoring.

3 Principles

3.1 *3D Freehand Ultrasound*

3D freehand US consists of acquiring 2D US images (B-scans) and simultaneously tracking the position and orientation of a probe in space with a position sensor or an optical or electromagnetic tracker system, as shown in Fig. 1. Thus, knowing the probe's position and the rigid-body transformation from the US image plane to the probe is sufficient for locating the B-scans in 3D space. The global coordinate system is usually the tracker frame. We can estimate the transformation between the image plane and the probe through a calibration procedure using a phantom, i.e., an object of known geometry. Many phantoms have been proposed in the literature, but the simplest is a point target[33] scanned from different positions and orientations. We can build it with cross-wires, a spherical object, or the tip of a stylus. The point phantom of two cross-wires is easy to build and produces accurate calibration results.[34–36]

3.2 *Structured Light*

Optical 3D shape measurement based on SL is a well-known optical metrology technique with many applications.[7] Its basic configuration is that of a camera–projector pair, as shown in Fig. 2. The projector projects a structured pattern (typically a sinusoidal pattern) onto the object's surface. It is distorted due to the topographic variations (indicated by a red line). This pattern is captured with a camera and processed to recover the depth information. Phase-shifting is one of the most used phase-retrieval algorithms since it produces pixelwise phase-maps using at least three patterns.[37] The obtained phase map is wrapped in the range from $-\pi$ to π. Therefore, an unwrapping procedure is necessary to estimate a continuous phase-map. Finally, with an unwrapped phase distribution, the projector–camera pixel correspondence is solved to perform triangulation and obtain the 3D surface.[38] This approach assumes that the SL system has been previously calibrated as a stereo vision system, in which the projector is modeled as an inverse camera.[39]
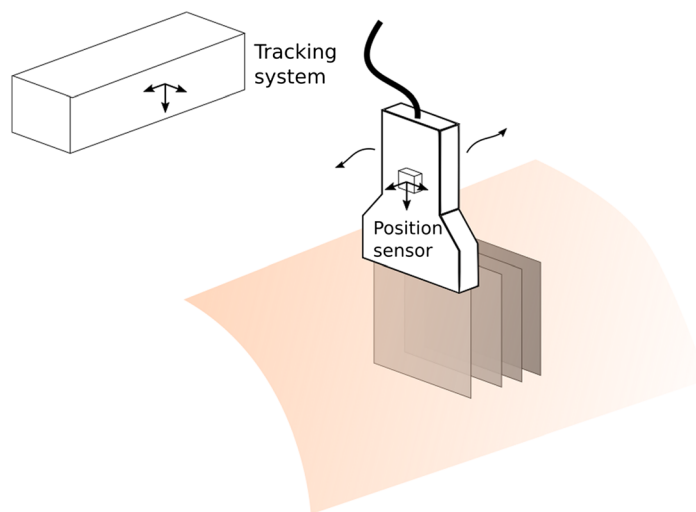

Fig. 1 3D freehand US technique: a tracking system is used to store the probe's pose along with the B-scan. As a result, the 2D US images can be mapped to a 3D volume.
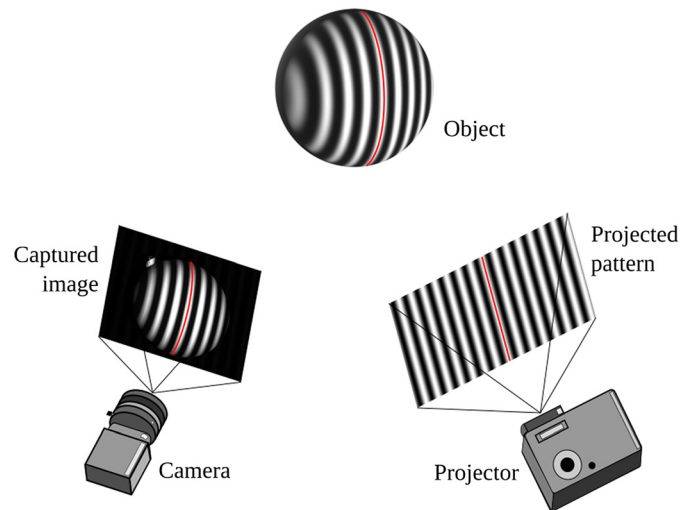
Fig. 2 An SL setup composed of a camera–projector pair. The projected pattern is distorted due to the object's topography. The red line shows how the pattern is deformed and how the camera detects it.

4 Method

Our multimodal system shown in Fig. 3 consists of two monochromatic Basler acA1300-200um cameras (1280×1024, 203 fps), a DLP Dell M318WL projector (1280×800), and a B-mode Biocare iS 20 US machine. The figure shows the main coordinate frames involved in the proposed multimodal imaging method. The world coordinate system $\{W\}$ is placed on the camera 1 frame $\{Cam_1\}$, and the remaining components are referred to this system. The pair camera-1 and projector form the SL system, whereas camera-1 and 2 form the stereo vision system used for tracking the US probe, which is the basis of the freehand US system. It is worth noting that, through the acquisition of both techniques, we do not change the position of the $\{W\}$ frame. Therefore, we avoid registering images for merging the multimodality data as both 3D data are referenced to the same coordinate system.

The proposed 3D multimodal imaging pipeline is described in Fig. 4. On the one hand, to obtain the 3D surface with the SL system, the projector projects vertical patterns onto the object's surface, which camera 1 captures. Using temporal phase unwrapping with the phase-shifting + gray coding technique, we recover the absolute phase map. With the absolute phase map, the camera–projector matches are established, and the object's topography is obtained through triangulation. On the other hand, for the 3D freehand US system, we use a strategy for tracking the probe's pose based on deep learning and using the stereo vision system $\{Cam_1\}$ and $\{Cam_2\}$ shown in Fig. 3. With in-house developed acquisition software, we simultaneously acquire the US and camera images.
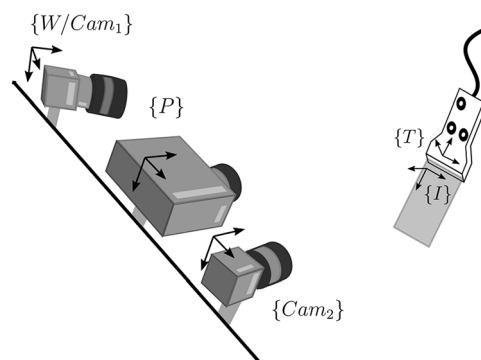

Fig. 3 Multimodal 3D imaging system composed of two cameras, a projector, and a US machine.
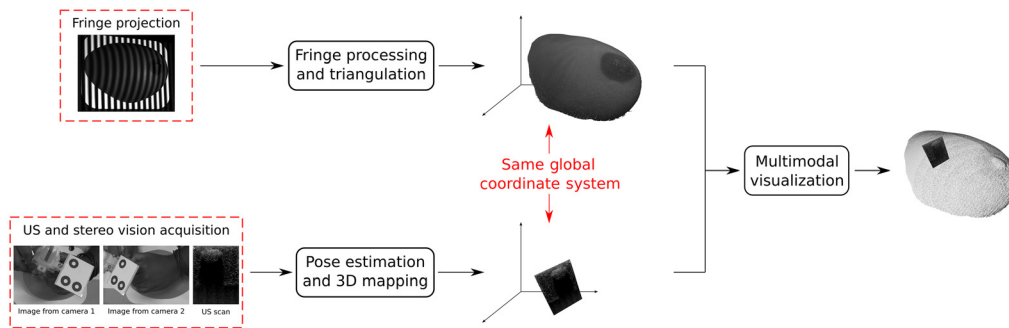
Fig. 4 The proposed multimodal 3D imaging pipeline with the external 3D surface and internal US images.

For tracking, a target with three coplanar circles is attached to the probe with a 3D-printed piece, as shown in Fig. 4. Then, with a convolutional neural network (CNN) model, we estimate the three circles' centers with subpixel resolution. The network also solves the matching problem between the two views. With the previous data, we estimate the position and orientation of the US probe. Afterward, the acquired US images are mapped to a 3D space by knowing the transformation between the US image frame $\{I\}$ and the target frame $\{T\}$, which is established through a calibration procedure. Finally, both 3D reconstructions are visualized in the same coordinate frame.

4.1 *Pose Estimation Procedure*

Recent freehand US systems use dedicated costly optical tracking systems that cannot be easily merged with complementary imaging techniques due to proprietary software and calibration limitations.[5,40] Alternative optical tracking methods with fiducial markers typically work well under controlled environments. In a previous version of this work,[41] we developed a target tracking method based on OpenCV using classical computer vision techniques. Although this method yields accurate tracking and pose estimation results, it often required tuning many parameters to make it work under different environments. For this reason, in this paper, we use MarkerPose,[42] a low-cost, real-time pose estimation method based on stereo vision and deep learning.

Our three-circles target and their ID labels are used for pose estimation, as shown in Fig. 5. With the centers' 3D position of the three coplanar circles, we define a coordinate system, where c_0 is the origin and the vectors $\overrightarrow{c_0c_1}$ and $\overrightarrow{c_0c_2}$ represent the x axis and y axis, respectively. However, the pose estimation requires detecting these points in subpixel accuracy and solving their correspondences from the two views.

The pose estimation of the target is addressed in three stages using two CNNs. The first two stages consists of the subpixel centers detection of the target as shown Fig. 6. In the first stage, the centers are detected in a pixel-level accuracy, using a SuperPoint-like network proposed by Hu et al.[43] This network is a two-headed network based on the SuperPoint architecture[44] in
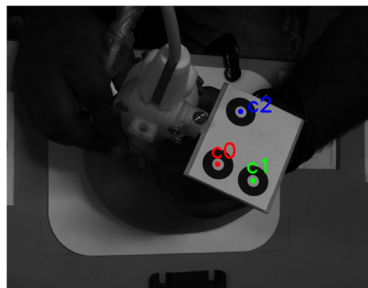

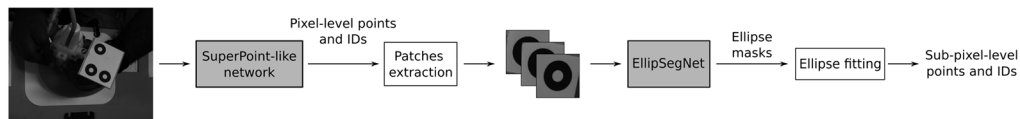**Fig. 5** The target with the corresponding circle centers' IDs.

Fig. 6 Subpixel circle centers detection and IDs classification pipeline. The gray-shaded boxes correspond to CNNs.
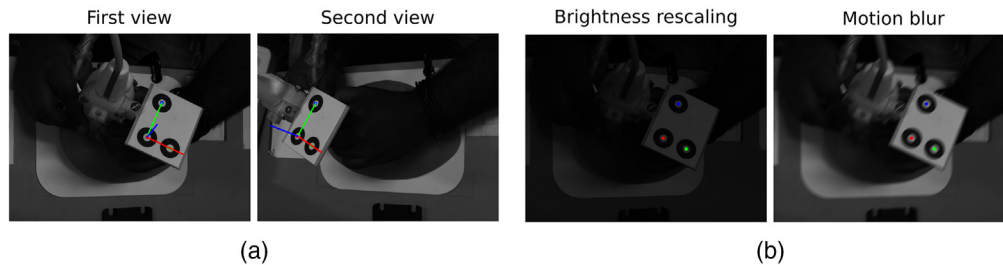

Fig. 7 Example of the robustness of the target tracking method. (a) Target detection under synthetically degraded lighting and motion blur. (b) Final pose estimation with the stereo images.

which an encoder is used to reduce the dimensionality of the input image. After the encoder, one of the heads estimates the probability that a circle center is present in each pixel (i.e., pixel-level detection). The other head classifies each point with a specific ID, solving the correspondence between the two views. In the second stage, with these rough detections, the algorithm extracts three patches centered on each circle center, such that the black circle is within the patch. As the circles project as ellipses, the detected contours are segmented with EllipSegNet, an encoder–decoder network. With the contour, the final subpixel center is estimated through ellipse fitting. Although it is known that the center of a projected circle and an ellipse deviate, for a small circle, this deviation is negligible.[45] In the last stage, the 3D coordinates of the three points are estimated through triangulation. With these points, the position and orientation of the target are calculated with respect to the $\{W\}$ frame. The translation vector $\mathbf{t}$ corresponds to the 3D coordinates of the center c_0. The rotation matrix can be established with the unit vectors of the target frame: $\hat{\mathbf{x}}$ calculated with c_0 and c_1 3D centers and $\hat{\mathbf{y}}$ calculated with c_0 and c_2 points. In this way, the rotation matrix is $\mathbf{R} = [\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{x}} \times \hat{\mathbf{y}}]$. Figure 7(a) shows an example of the acquired stereo images of the target and the final estimated target pose.

Both center detection networks were trained with real images. For the SuperPoint variant, a total of 5683 grayscale images were used. For the EllipSegNet, a total of 11,010 patches were used for training. The center detection of the first two stages is robust to low lighting conditions, which is important for using the DLP projector in the SL system. We have to set a trade-off between the camera aperture and the exposure to avoid overexposed images for the SL system. This trade-off often leads to suboptimal images for freehand US probe tracking, and the aperture cannot be modified to ensure that the calibration remains the same. The tracking procedure is also robust to motion blur, which is crucial for real-time tracking of the US probe. Figure 7(b) shows an example of the detection under synthetic low lighting and with synthetic motion blur applied. It is worth noting that it was trained and validated with more severe low-lighting and motion blur.[42]

4.2 *3D Freehand Ultrasound Calibration*

Calibration of a 3D freehand US system requires tracking the probe's pose and consists of estimating the geometric relationship between the US image plane and the probe. For the calibration procedure, we used a point phantom of two cross-wires. In Fig. 8, we illustrate the spatial relationships of the five coordinate systems involved in the probe calibration with our stereo vision system.
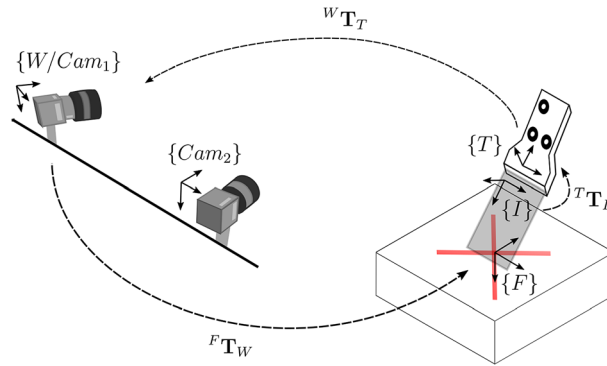
Fig. 8 Transducer calibration: geometric relations and reference frames.

Using the transformation matrix $^W\mathbf{T}_T$ from the transducer's coordinate system $\{T\}$ to the world coordinate frame $\{W\}$ given by the stereo vision system, we need to estimate the transformation $^T\mathbf{T}_I$ from the image frame $\{I\}$ to the transducer's frame $\{T\}$. Furthermore, we also need to calculate the x and y scales of the US image s_x and s_y in millimeters per pixel to convert a pixel (u, v) of the B-scan to metric units. The phantom coordinate system $\{F\}$ is placed so that its origin coincides with the cross-wires point. In this way, if we acquire a B-scan of the cross-wire phantom, we can map the segmented image coordinate of the point target in pixels to the physical phantom frame $\{F\}$ as

$$[0\ 0\ 0\ 1]^T = {}^F\mathbf{T}_W{}^W\mathbf{T}_T{}^T\mathbf{T}_I[s_x u\ s_y v\ 0\ 1]^T. \tag{1}$$

With this equation and a total of N US images of the phantom, we can estimate the unknowns involved in the calibration.

To evaluate the transducer calibration quality, we measure the precision with the calibration reproducibility (CR) using the methods proposed by Prager et al.[34] and Hsu et al.[46] A specific point (u, v) in the B-scan is reconstructed from the image plane to the transducer coordinate system, and in this frame, we measure the variability of this point. Prager et al.[34] used the bottom right pixel (u_{max}, v_{max}) of the B-scan to measure CR under two different calibrations with the following expression:

$$\mu_{CR1} = |{}^T\mathbf{T}_{I1}\mathbf{X}_1^I - {}^T\mathbf{T}_{I2}\mathbf{X}_2^I|, \tag{2}$$

where $\mathbf{X}_i^I = [s_{xi}u_{max}, s_{yi}v_{max}, 1, 0]^\mathsf{T}$. We calculate CR with this expression using all possible pairs of calibration parameters, and finally, the mean of these values is reported as the final precision.

With the proposed strategy by Hsu et al.,[46] CR is assessed as

$$\mu_{CR2} = \frac{1}{N}\sum_{i=1}^{N}|{}^T\mathbf{T}_{Ii}\mathbf{X}_i^I - \overline{\mathbf{X}}^T|, \tag{3}$$

where $\overline{\mathbf{X}}^T$ is the mean in each dimension of all of the reconstructed points $\mathbf{X}_i^I$ in the transducer $\{T\}$ coordinate system. As trial points, they used the center and the four corners of the image.

4.3 *Structured Light and Stereo Vision Calibration*

As our system is composed of two cameras and a projector, we need to calibrate the SL system for 3D reconstruction and the stereo vision system for tracking the US probe. As we can regard the projector as an inverse camera,[39] we calibrate both systems with the well-known stereo calibration procedure using a pinhole camera model.[47] The $\{Cam_1\}$-$\{P\}$ and $\{Cam_1\}$-$\{Cam_2\}$ pairs are calibrated independently using the same pattern of asymmetric circles. For establishing the camera–projector correspondences, we used the phase-aided method proposed by Zhang and Huang[39] projecting fringe patterns onto the calibration board. To obtain a reliable phase-aided

correspondence, we used binary defocused fringes of an 18-pixel pitch with the 18-step phase-shifting algorithm for phase recovery along with 7-bit gray coding patterns for absolute phase unwrapping.[48] To ensure high accuracy, we acquired images from 34 poses of the calibration target.

After obtaining the $\{Cam_1\}$-$\{P\}$ and $\{Cam_1\}$-$\{Cam_2\}$ point correspondences, the calibration procedure is the same for both systems. We briefly describe the procedure for the SL system. The calibration consists of estimating the intrinsic and extrinsic parameters of the camera and the projector using the pinhole camera model for both devices, given as

$$s\mathbf{x} = \mathbf{KMX}, \tag{4}$$

where $\mathbf{X}$ is a point in the 3D world coordinate system, $\mathbf{x}$ is its projection in the camera or projector sensor, $\mathbf{K}$ is the intrinsic parameter matrix, and $\mathbf{M}$ is the extrinsic parameter matrix. In its expanded form, Eq. (4) becomes

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \tag{5}$$

where f_u and f_v are the focal length in the u and v directions, respectively, (c_u, c_v) is the principal point, r_{ij} are the rotation matrix parameters, and t_i are the translation coefficients between the camera or projector and the world frame. As $\{Cam_1\}$ is also the world frame, the camera rotation matrix is the identity matrix, and the translation vector is the zero vector. Thus, for the camera and the projector, we have two sets of equations given by

$$s^c\mathbf{x}^c = \mathbf{K}^c\mathbf{M}^c\mathbf{X}^w, \tag{6}$$

$$s^p\mathbf{x}^p = \mathbf{K}^p\mathbf{M}^p\mathbf{X}^w. \tag{7}$$

With different views of the calibration pattern, we use Eqs. (6) and (7) to solve for the camera's and projector's intrinsic parameters. Finally, we estimate the extrinsic parameters between the camera–projector pair. An analogous procedure is carried out for calibrating the stereo vision system formed by the two cameras. The obtained reprojection errors for camera-1, projector, and camera-2 are 0.1384, 0.1312, and 0.1508, respectively, which are quite small.

For 3D reconstruction with the SL system, we used an eight-step phase-shifting method with a center-line image for absolute phase-unwrapping to solve the phase correspondence. Equations (6) and (7) are solved to obtain the 3D surface.

5 Experiments and Results

We carried out different experiments to evaluate both the 3D freehand US and the SL systems separately and jointly as a multimodal technique.

5.1 *3D Freehand US Calibration Assessment*

With the $\{Cam_1\} - \{Cam_2\}$ system calibrated, we carried out a total of five calibrations of the US probe with 30 images in each one, using a linear transducer set to a 5 cm depth. Table 1 shows the results of the calibration precision evaluated with μ_{CR1} and μ_{CR2}, using Eqs. (2) and (3), respectively. We use five trial points for the CR evaluation for both metrics, the center and the four image corners, where our US image is 321×408 px. Our results are similar to those previously reported. For example, HS et al.[46] report a CR of 0.27 mm with μ_{CR2}, measured in the center of the image and using a point phantom and the probe at 3 cm depth. Furthermore, Lindseth et al.[49] report a CR at the center of the B-scan of 0.62 mm with a point phantom and

Table 1 Precision assessment: CR results with a total of five calibrations.

Trial point	μ_{CR1} (mm)	μ_{CR2} (mm)
Center	0.4337	0.2736
Bottom right (u_{max}, v_{max})	0.5513	0.3453
Mean (center and four corners)	0.5905	0.3681

Table 2 3D reconstruction assessment results under different poses of a flat board.

Pose	1	2	3	4	5
RMS error (mm)	0.1161	0.1244	0.1305	0.1120	0.1006
x range (mm)	[−116, 126.07]	[−120.58, 129.91]	[−116.4, 132.39]	[−119.09, 130.12]	[−116.26, 129.52]
y range (mm)	[−96.75, 90.41]	[−99.97, 93.57]	[−101.23, 97.33]	[−99.85, 88.38]	[−99.39, 84.78]
z range (mm)	[659.89, 680.09]	[682.8, 703.48]	[675.58, 713.57]	[642.23, 701.56]	[606.16, 697.98]

a linear probe at 8 cm depth using μ_{CR1}. Finally, we report an RMS error of all equations obtained through the five calibrations of 0.4231 mm.

5.2 Structured Light Assessment

To evaluate the calibration of the SL system, we reconstructed a plane in five different poses. The volume covered with the planes is $252.97 \times 198.56 \times 107.41$ mm. The RMS errors between the reconstructed and ideal planes estimated through least-squares is shown in Table 2.

We also measured a sphere with a 19.8-mm radius to evaluate the system calibration with a different geometry. The estimated radius with the 3D reconstruction is 19.755 mm, which gives an absolute error of 0.045 mm. Finally, the RMS error between the reconstructed and ideal spheres is 0.0399 mm.

5.3 Multimodal System Assessment

We evaluated the proposed system as a multimodal technique through two experiments. For a quantitative evaluation, we designed an experiment in which we measured two concentric cylinders. Figure 9(a) shows a schematic of the experiment, in which the inner cylinder is submerged in water to be measured with 3D freehand US, while the outer cylinder is reconstructed with SL. This experiment aims to estimate the distance d between the inner and outer cylinders, i.e., between the 3D freehand US reconstruction and the SL reconstruction. Figure 9(b) shows
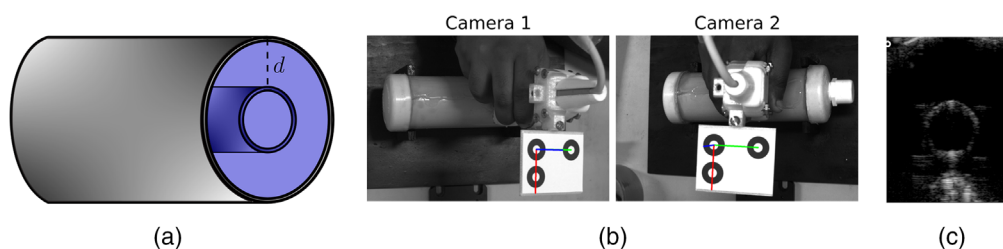

Fig. 9 Validation experiment of the multimodal system with two concentric cylinders. (a) Diagram of the arrangement of cylinders and the measured distance d between the internal and external reconstructions. (b) Captured stereo images during US acquisition of the internal cylinder. (c) B-scan example of the internal cylinder.
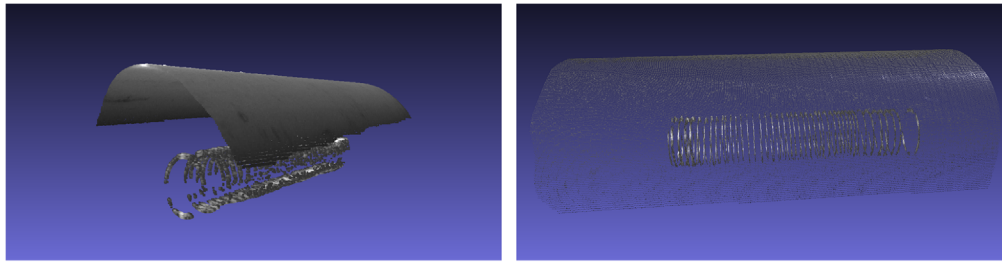
Fig. 10 Internal and external reconstruction results of the concentric cylinders in the same coordinate system used for the validation experiment.
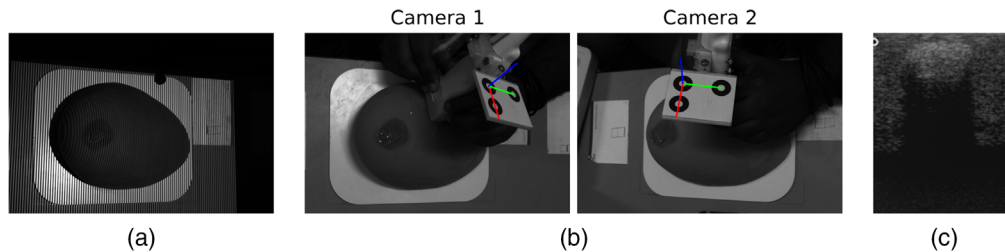

(a)   (b)   (c)

Fig. 11 SL and freehand US acquisition of the phantom for the qualitative experiment. (a) Projected pattern acquisition. (b) Stereo images with estimated target pose for freehand US. (c) US image.

the stereo images of the measured object during US acquisition, and Fig. 9(b) the corresponding US image. The outer diameter of the internal cylinder is 15.82 mm, that of the external cylinder is 60 mm, and the distance d between the cylinders is 22.38 mm.

For this experiment, a total of 94 US slices were captured along the cylinder. The multimodal reconstruction results are shown in Fig. 10, in which we have the segmented rings from the US images mapped to the 3D space and the external cylindrical cap reconstructed with the SL system. With the point cloud obtained with each technique, we estimated the least-squares ideal cylinder. The internal cylinder's estimated diameter with the freehand US reconstruction is 15.1 mm, which gives an absolute error of 0.72 mm. The external diameter of the outer cylinder is 60.10 mm with 0.1 mm of absolute error. Finally, the estimated d distance between both reconstructions is 22.50 mm with an absolute error of 0.12 mm. These results show an adequate performance of the proposed multimodal 3D imaging technique.

The second experiment consisted of evaluating the proposed multimodal system using a breast phantom 3B SONOtrain P125 made from US material with three tumors. $170 \times 130 \times 55$ mm are the phantom's dimensions, and the tumors have three different sizes: 27, 14, and 12 mm in diameter. We reconstructed the external breast surface (external features) with the SL technique and the three tumors (internal features) with the 3D freehand US method. Thus, we projected the SL patterns for surface 3D reconstruction as shown Fig. 11(a). Then, we acquired US slices of the three tumors for the 3D mapping. An example of the pose estimation with the stereo vision system is shown in Fig. 11(b), and the B-scan image simultaneously acquired with the stereo images is shown in Fig. 11(c). Figure 12 shows the result of both reconstructions
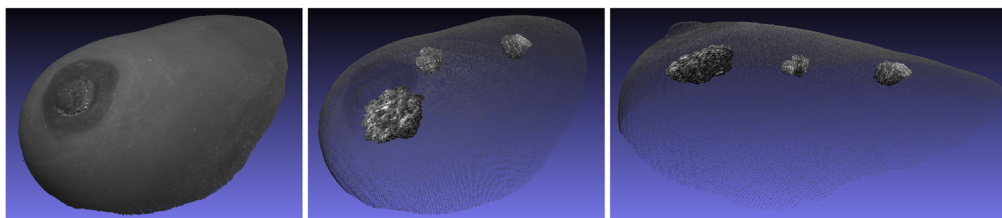

Fig. 12 Qualitative results of the multimodal reconstruction. The 3D surface of the breast phantom from the SL technique, and the three tumors of the phantom reconstructed with 3D freehand US.
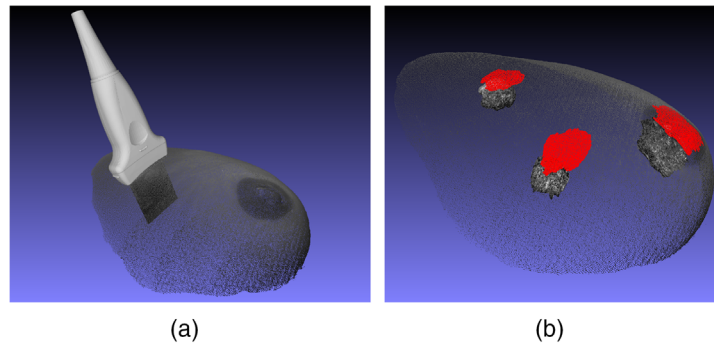
Fig. 13 Potential applications of the proposed 3D multimodal imaging system. (a) The proposed multimodal technique as a navigation system. The tracked US probe mapped in 3D with the US scan and the 3D surface. (b) The projector as an active device to project the internal structures in the external surface.

relative to the $\{W\}$. Both reconstructions are in the same coordinate system, and the reconstructed tumors are totally within the phantom surface. The result is in agreement with the expected position of the tumors.

6 Potential Applications of the Multimodal System

As we track the US probe's pose in the same coordinate system of the 3D reconstructions, we can use the proposed multimodal system as a navigation system in which the probe can be displayed in real-time with the reconstructions as shown in Fig. 13(a). In this way, we have the US slice mapped in 3D and the probe with the 3D surface reconstruction. This visualization can be useful for image interpretation and spatial understanding during surgery. In addition, we can track any other instrument by attaching a marker to display it in 3D with the other 3D reconstructions, e.g., a needle for biopsy procedures.

Furthermore, the DLP projector can be used as an active device to project the internal reconstructed structures onto the subject's skin. In a similar fashion to fluorescence-guided surgery, the internal structures are highlighted by projecting light onto the tissue. For example, in breast cancer biopsy.[50] We can also display with the projector the location of the internal 3D structures. Figure 13(b) shows a simulated example with the reconstructed tumors highlighted on the breast's surface.

7 Summary

We proposed a low-cost 3D multimodal medical imaging technique by combining 3D freehand US and SL. With this multimodal technique, it is possible to complement the internal information acquired with 3D freehand US with the external surface obtained from SL. Both reconstructions are in the same global coordinate system, avoiding complicated data registration procedures. Furthermore, using a pose estimation system based on CNNs and stereo vision avoids using costly commercial tracking systems. The experimental results show the proposed system's high accuracy and its suitability for preoperative or intraoperative procedures such as surgical planning or guidance.

Acknowledgments

Code, Data, and Materials Availability

The software implementation is available at https://github.com/jhacsonmeza/StructuredLight_3DfreehandUS. The image acquisition software was developed in C++, and the proposed method for optical tracking and SL 3D reconstruction was implemented in Python using OpenCV 4.4.0. The deep learning model was implemented with PyTorch 1.6.0. In the current implementation, the data were acquired and processed offline for visualization. A real-time implementation is currently being developed.

References

1. P. Mascagni et al., "New intraoperative imaging technologies: innovating the surgeon's eye toward surgical precision," *J. Surg. Oncol.* **118**, 265–282 (2018).
2. E. J. R. van Beek et al., "Value of MRI in medicine: more than just another test?" *J. Magn. Reson. Imaging* **49**, e14–e25 (2019).
3. S. H. C. Ortiz, T. Chiu, and M. D. Fox, "Ultrasound image enhancement: a review," *Biomed. Signal Process. Control* **7**(5), 419–428 (2012).
4. Q. Huang and Z. Zeng, "A review on real-time 3d ultrasound imaging technology," *Biomed. Res. Int.* **2017**, 6027029 (2017).
5. E. Colley et al., "A methodology for non-invasive 3-d surveillance of arteriovenous fistulae using freehand ultrasound," *IEEE Trans. Biomed. Eng.* **65**(8), 1885–1891 (2018).
6. S. Zhang, "High-speed 3D shape measurement with structured light methods: a review," *Opt. Lasers Eng.* **106**, 119–131 (2018).
7. A. G. Marrugo, F. Gao, and S. Zhang, "State-of-the-art active optical techniques for three-dimensional surface metrology: a review [Invited]," *J. Opt. Soc. Am. A* **37**(9), B60–18 (2020).
8. F. Zhang et al., "Coaxial projective imaging system for surgical navigation and telementoring," *J. Biomed. Opt.* **24**, 105002 (2019).
9. S. Van der Jeught and J. J. J. Dirckx, "Real-time structured light-based otoscopy for quantitative measurement of eardrum deformation," *J. Biomed. Opt.* **22**, 016008 (2017).
10. T. T. Quang et al., "Fluorescence imaging topography scanning system for intraoperative multimodal imaging," *PLoS One* **12**, e0174928 (2017).
11. E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Med. Image Anal.* **48**, 107–116 (2018).
12. M. Anzidei et al., "Imaging-guided chest biopsies: techniques and clinical results," *Insights Imaging* **8**(4), 419–428 (2017).
13. A. K. Bowden et al., "Optical technologies for improving healthcare in low-resource settings: introduction to the feature issue," *Biomed. Opt. Express* **11**, 3091–3094 (2020).
14. S. R. Cherry, "Multimodality imaging: beyond PET/CT and SPECT/CT," *Semin. Nucl. Med.* **39**(5), 348–353 (2009).
15. T. L. Walker, R. Bamford, and M. Finch-Jones, "Intraoperative ultrasound for the colorectal surgeon: current trends and barriers," *ANZ J. Surg.* **87**(9), 671–676 (2017).
16. B. Li et al., "Ultrasound guided fluorescence molecular tomography with improved quantification by an attenuation compensated born-normalization and *in vivo* preclinical study of cancer," *Rev. Sci. Instrum.* **85**(5), 053703 (2014).
17. T. A. N. Hernes et al., "Computer-assisted 3d ultrasound-guided neurosurgery: technological contributions, including multimodal registration and advanced display, demonstrating future perspectives," *Int. J. Med. Rob. Comput. Assisted Surg.* **2**(1), 45–59 (2006).
18. F. Lindseth et al., "Multimodal image fusion in ultrasound-based neuronavigation: improving overview and interpretation by integrating preoperative MRI with intraoperative 3d ultrasound," *Comput. Aided Surg.* **8**(2), 49–69 (2003).
19. H. Fatakdawala et al., "Multimodal *in vivo* imaging of oral cancer using fluorescence lifetime, photoacoustic and ultrasound techniques," *Biomed. Opt. Express* **4**(9), 1724–1741 (2013).
20. Y. Li, J. Chen, and Z. Chen, "Multimodal intravascular imaging technology for characterization of atherosclerosis," *J. Innovative Opt. Health Sci.* **13**(1), 2030001 (2020).

21. C. Mela, F. Papay, and Y. Liu, "Novel multimodal, multiscale imaging system with augmented reality," *Diagnostics* **11**(3), 441 (2021).

22. D. M. McClatchy, III et al., "Calibration and analysis of a multimodal micro-CT and structured light imaging system for the evaluation of excised breast tissue," *Phys. Med. Biol.* **62**(23), 8983 (2017).

23. O. V. Olesen et al., "Motion tracking for medical imaging: a nonvisible structured light tracking approach," *IEEE Trans. Med. Imaging* **31**(1), 79–87 (2012).

24. L. Pino-Almero et al., "Quantification of topographic changes in the surface of back of young patients monitored for idiopathic scoliosis: correlation with radiographic variables," *J. Biomed. Opt.* **21**(11), 116001 (2016).

25. C.-W. J. Cheung et al., "Freehand three-dimensional ultrasound system for assessment of scoliosis," *J. Orthop. Translat.* **3**(3), 123–133 (2015).

26. R. Vairavan et al., "A brief on breast carcinoma and deliberation on current non-invasive imaging techniques for detection," *Curr. Med. Imaging Rev.* **13**, 85–121 (2017).

27. F. Šroubek et al., "A computer-assisted system for handheld whole-breast ultrasonography," *Int. J. Comput. Assist. Radiol. Surg.* **14**(3), 509–516 (2019).

28. W. Norhaimi et al., "Breast surface variation phase map analysis with digital fringe projection," *Proc. SPIE* **11197**, 1119717 (2019).

29. S. Horvath et al., "Towards an ultrasound probe with vision: structured light to determine surface orientation," *Lect. Notes Comput. Sci.* **7264**, 58–64 (2011).

30. E. Basafa et al., "Visual tracking for multi-modality computer-assisted image guidance," *Proc. SPIE* **10135**, 101352S (2017).

31. S.-Y. Sun, M. Gilbertson, and B. W. Anthony, "Probe localization for freehand 3d ultrasound by tracking skin features," *Lect. Notes Comput. Sci.* **8674**, 365–372 (2014).

32. J. Wang et al., "Ultrasound tracking using probesight: camera pose estimation relative to external anatomy by inverse rendering of a prior high-resolution 3d surface map," in *IEEE Winter Conf. Appl. Comput. Vision*, IEEE, pp. 825–833 (2017).

33. P.-W. Hsurager, et al., "Comparison of freehand 3-d ultrasound calibration techniques using a stylus," *Ultrasound Med. Biol.* **34**(10), 1610–1621 (2008).

34. R. W. Prager et al., "Rapid calibration for 3-d freehand ultrasound," *Ultrasound Med. Biol.* **24**(6), 855–869 (1998).

35. L. Mercier et al., "A review of calibration techniques for freehand 3-d ultrasound systems," *Ultrasound Med. Biol.* **31**(4), 449–471 (2005).

36. F. Torres et al., "Image tracking and volume reconstruction of medical ultrasound," *Rev. mexicana ingeniera biomed.* **33**(2), 101–115 (2012).

37. L. Lu et al., "Motion induced error reduction methods for phase shifting profilometry: a review," *Opt. Lasers Eng.* **141**, 106573 (2021).

38. R. Juarez-Salazar et al., "Key concepts for phase-to-coordinate conversion in fringe projection systems," *Appl. Opt.* **58**(18), 4828–4834 (2019).

39. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," *Opt. Eng.* **45**(8), 083601 (2006).

40. Y. Hu et al., "Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks," *Lect. Notes Comput. Sci.* **10555**, 105–115 (2017).

41. J. Meza et al., "A low-cost multi-modal medical imaging system with fringe projection profilometry and 3D freehand ultrasound," *Proc. SPIE* **11330**, 1133004 (2020).

42. J. Meza, L. A. Romero, and A. G. Marrugo, "Markerpose: robust real-time planar target tracking for accurate stereo pose estimation," https://arxiv.org/abs/2105.00368 (2021).

43. D. Hu, D. DeTone, and T. Malisiewicz, "Deep ChArUco: dark ChArUco marker pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 8436–8444 (2019).

44. D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: self-supervised interest point detection and description," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 224–236 (2018).

45. Y. Sun, "Analysis for center deviation of circular target under perspective projection," *Eng. Comput.* **36**(7), 2403–2413 (2019).

46. P.-W. Hsu et al., "Freehand 3D ultrasound calibration: a review," in *Advanced Imaging in Biology and Medicine*, C. W. Sensen and B. Hallgrímsson, eds., pp. 47–84, Springer, Berlin, Heidelberg (2009).

47. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).

48. S. Zhang, *High-Speed 3D Imaging with Digital Fringe Projection Techniques*, CRC Press (2016).

49. F. Lindseth et al., "Probe calibration for freehand 3-d ultrasound," *Ultrasound Med. Biol.* **29**(11), 1607–1623 (2003).

50. B. E. Schaafsma et al., "Clinical trial of combined radio- and fluorescence-guided sentinel lymph node biopsy in breast cancer," *Br. J. Surg.* **100**(8), 1037 (2013).

Jhacson Meza is a graduate student at the Universidad Tecnológica de Bolívar (UTB), Cartagena, Colombia, where he received his BEng degree in mechatronics engineering. He received the Excellence graduate scholarship from UTB and a "Joven Talento" scholarship from MinCiencias. His current research interests include computer vision and machine learning.

Sonia H. Contreras-Ortiz received her BE degree in electronic engineering and her MS degree in electrical power from the Universidad Industrial de Santander in Bucaramanga, Colombia, and received her PhD in biomedical engineering from the University of Connecticut, in Storrs, Connecticut, USA, in 2011. Currently, she is the director of the biomedical engineering program at UTB, in Cartagena de Indias, Colombia. Her research interests include biosignal and medical image processing and analysis. She is a member of SPIE.

Lenny A. Romero received her BSc and MSc degree in physics from the Universidad Industrial de Santander, Colombia, and her PhD in optical engineering (cum laude) from the Universitat Politècnica de Catalunya, Spain. Currently, she is the dean of the Faculty of Basic Sciences at the UTB, Colombia. Her research interests are optical metrology, visual optics, and 3D imaging for biomedical applications. She is a member of SPIE.

Andres G. Marrugo is a professor of mechatronics engineering at the UTB, Colombia, where he obtained his BEng degree in mechatronics engineering (summa cum laude). He received his PhD in optical engineering (cum laude) from the Technical University of Catalonia, Barcelona, Spain. He received the 2019 SPIE DCS Rising Researcher Award and the 2019 Fulbright Visiting Scholar Award. His research interests include optical metrology, 3D computer vision, and biomedical image analysis. He is a member of SPIE.