2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10341168

Robust 3D breast reconstruction based on monocular images and artificial intelligence for robotic guided oncological interventions

Bruno Duarte, Bruno Oliveira, Helena R. Torres, Pedro Morais, Jaime C. Fonseca, and João L. Vilaça

Abstract— **Breast cancer is a global public health concern. For women with suspicious breast lesions, the current diagnosis requires a biopsy, which is usually guided by ultrasound (US). However, this process is challenging due to the low quality of the US image and the complexity of dealing with the US probe and the surgical needle simultaneously, making it largely reliant on the surgeon's expertise. Some previous works employing collaborative robots emerged to improve the precision of biopsy interventions, providing an easier, safer, and more ergonomic procedure. However, for these equipment to be able to navigate around the breast autonomously, 3D breast reconstruction needs to be available. The accuracy of these systems still needs to improve, with the 3D reconstruction of the breast being one of the biggest focuses of errors. The main objective of this work is to develop a method to obtain a robust 3D reconstruction of the patient's breast, based on RGB monocular images, which later can be used to compute the robot's trajectories for the biopsy. To this end, depth estimation techniques will be developed, based on a deep learning architecture constituted by a CNN, LSTM, and MLP, to generate depth maps capable of being converted into point clouds. After merging several from multiple points of view, it is possible to generate a real-time reconstruction of the breast as a mesh. The development and validation of our method was performed using a previously described synthetic dataset. Hence, this procedure takes RGB images and the cameras' position and outputs the breasts' meshes. It has a mean error of 3.9 mm and a standard deviation of 1.2 mm. The final results attest to the ability of this methodology to predict the breast's shape and size using monocular images.**

Clinical Relevance— **This work proposes a method based on artificial intelligence and monocular RGB images to obtain the breast's volume during robotic guided breast biopsies, improving their execution and safety.**

I. Introduction

Breast cancer is one of the most prevalent cancers worldwide and one of the most dreadful diseases. In 2020, 2.3 million new cases of female breast cancer were detected [1] [2]. Breast cancer cells can be situated in the milk ducts, lobules, lymph vessels, and/or breast tissue, which determines the type of breast cancer [3]. Breast cancer can be detected when an abnormal lump is found or when some changes are identified on the nipple or breast skin, however, generally, it is diagnosed during routine screenings, where mammography is the most common procedure [4]. The mammogram is analyzed and the lesion is classified within one of the BIRADS categories [5]. If the result is 4 or 5, there is a suspicion that the tumor might be malign, which demands a breast biopsy to clarify it [5]. It can be performed under different guidance, but the US-guided breast biopsy is the preferred strategy because it is the cheapest method, allows real-time visualization of the needle, offers multi-directional sampling, and its US probe is facilely manipulable, making it easier to access every part of the breast and axilla. Additionally, it also avoids ionizing radiation exposure and intravenous contrast. However, the US-guided breast biopsy presents some technical execution issues, such as challenging techniques to perform the needle insertion, especially if the lesion is located in a hard-to-reach region, or when the US images present poor quality and/or artifacts [6]. Therefore, this procedure is widely dependent on the surgeon's expertise and capacity, requiring a lot of training for new professionals, which can take a lot of time and investment. Given the complexity of the procedure and the doctor's exhaustion associated with it, some biopsies end up being very time-consuming, which can lead to longer surgeries, less capacity to see multiple patients, and more biopsy errors.

These issues justified the emergence of new methods that integrate AI and Robotics to improve US-guided breast biopsy. One of the main challenges for these strategies is to obtain the breast's volume in real-time in order to control the robot and perform the US scanning around the breast. This allows the creation of a 3D US scan of the breast interior, to localize the lesion, and then precisely insert the biopsy needle [7]. Still, current reconstruction techniques require better precision and technologies that do not involve occlusions, expensive equipment, or minimal distances [7][8].

This work proposes an innovative Deep Learning (DL) architecture that fuses a CNN, LSTM, and MLP to obtain depth maps from monocular RGB images, which are then

This project was funded by the projects "NORTE-01-0145-FEDER-000045" and "NORTE-01-0145-FEDER-000059", supported by Northern Portugal Regional Operational Programme (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). It was also financed by national funds, through FCT - Foundation for Science and Technology and FCT / MCTES under the project UIDB / 05549/2020, UIDP/05549/2020, CEECINST/00039/2021 and LASI-LA/P/0104/2020. This project was also funded by the Innovation Pact HfFP Health From Portugal, co-funded from the "Mobilizing Agendas for Business Innovation" of the "Next Generation EU" program of Component 5 of the Recovery and Resilience Plan (RRP), concerning "Capitalization and Business Innovation", under the Regulation of the Incentive System "Agendas for Business Innovation".

The authors also acknowledge FCT, Portugal and the European Social Fund, European Union, for funding support through the "Programa Operacional Capital Humano" (POCH) in the scope of the PhD grants SFRH/BD/136721/2018 (B. Oliveira) and SFRH/BD/136670 (H. R. Torres).

B. Duarte is with 2Ai – School of Technology, IPCA, Barcelos, Portugal and with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (e-mail: bduarte@ipca.pt).

B. Oliveira and H. R. Torres are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal, with Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, and with ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal (e-mail:, boliveira@ipca.pt, htorres@ipca.pt).

P. Morais, and J. L. Vilaça are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, and LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal (e-mail: pmorais@ipca.pt, jvilaca@ipca.pt).

J. C. Fonseca is with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (e-mail: jaime@dei.uminho.pt).

converted to 3D meshes, in real-time. This methodology can benefit new robotic-guided breast biopsy approaches by improving their safety and execution and reducing their end-effector's cost.

II. Related Works

A. Robotic Guided Breast Biopsies

As technology advanced, some robotic devices were created and employed in breast biopsies. Most of the recent works are focused on MRI-guided breast biopsy, where the robot is encapsulated in the closed-bore MRI scanner [9][10]. Recently, an MRI and Ultrasound Robotic Assisted Biopsy (MURAB) project [11] was able to combine robotics and AI with MRI and US images. This work employs a structured light strategy to project patterns into the breasts and obtain their volume, which requires expensive hardware, namely a projector and a detector. With this volume, the robot can start acquiring US information by scanning the breast surface with a US probe. Consequently, a 3D US volume of the breast's interior is created and registered with the MRI data obtained before the biopsy. This is used to compute the robot coordinates for the needle insertion.

B. 3D Reconstruction

For the desired 3D reconstruction, several technologies can be considered. For instance, the Structure-from-Motion (SfM) technique can provide a 3D reconstruction of an object's geometry by overlapping images of this object captured from different points of view [13]. However, this procedure tends to generate distortions and exaggerate the smoothing effect. Nevertheless, this methodology is interesting, especially because a collaborative robot can be used to navigate through an initial path around the patient's breast to achieve this. Alternatively, there are also other works applying DL to 3D reconstruction, using 2D images as input [14]. Still, these approaches are complex and require a lot of pre-processing and time. The Ray-Onet[15] approach stands out because, besides RGB images, it also integrates the cameras' position in their network's input, which contributes to increasing the spatial information. The cameras' position is generally treated as tabular data, that can be treated by an MLP segment, as demonstrated by Ahsan et al work [16]. The depth estimation methods, based on DL, are easier to achieve and the predicted depth maps can be converted to 3D formats. The architecture proposed by the DenseDepth [17] work, follows a standard encoder-decoder strategy, with an interesting loss function, which can be a starting point for this development.

III. Methodology

This work utilized a toolchain, previously developed by the authors[12] to create a synthetic dataset based on synthetic human breasts. Then, a multi-input DL architecture, constituted by a CNN, LSTM, and a MLP segment was incrementally built. The synthetic dataset was used to train and test the DL model. The predicted depth maps for each human were converted to point clouds and the 7 most suitable were fused to generate merged point clouds, with more spatial information than the individual ones. Finally, these were converted to meshes, generating the 3D reconstruction of the breast. This 3D breast reconstruction pipeline is shown in Figure 1.

A. DL architecture for Depth Estimation

The DL architecture for depth estimation, in Figure 1, starts by pre-processing the synthetic data and then implementing a CNN, inspired by DenseDepth [17]. To further improve the results, an LSTM was introduced, since it strengthens the temporal information. An MLP was then added, enabling the feeding of the DL network with the camera's position, which increases the spatial information available. Thus, the implemented DL model receives two types of inputs: RGB images and the cameras' positions (x, y, z).

During pre-processing, the LSTM sequences were created and each image and depth map was transformed by applying a random horizontal flip and a random channel swap with a probability of 50% for data augmentation. The RGB images and depth maps were resized to 640x480 and 320x240, respectively. Then, the RGB images were divided by 255 and the depth maps were multiplied by the maximum depth, 100, converting them from meters to centimeters. A final normalization was applied to the depth maps using (1), inspired by the DenseDepth algorithm. This was employed to make the network give higher loss on sections closer to the camera.

$$\text{Normalized depth map} = \frac{\text{Maximum depth}}{\text{Depth map}} \qquad (1)$$

The implemented loss function, L, is shown in (2) where *pred* and *gt* refer to the predicted and ground truth depth maps, respectively, and θ refers to a weight parameter equal to 0.1. It is constituted by 3 loss functions: the SSIM loss, L_{SSIM}, the point-wise L_1 loss, L_{L1}, and the L_1 loss defined over the image gradient of the depth image, L_{Grad}. This loss function aims to reduce the discrepancy between the depth values while simultaneously penalizing distortions of high-frequency details in the depth map's image domain.

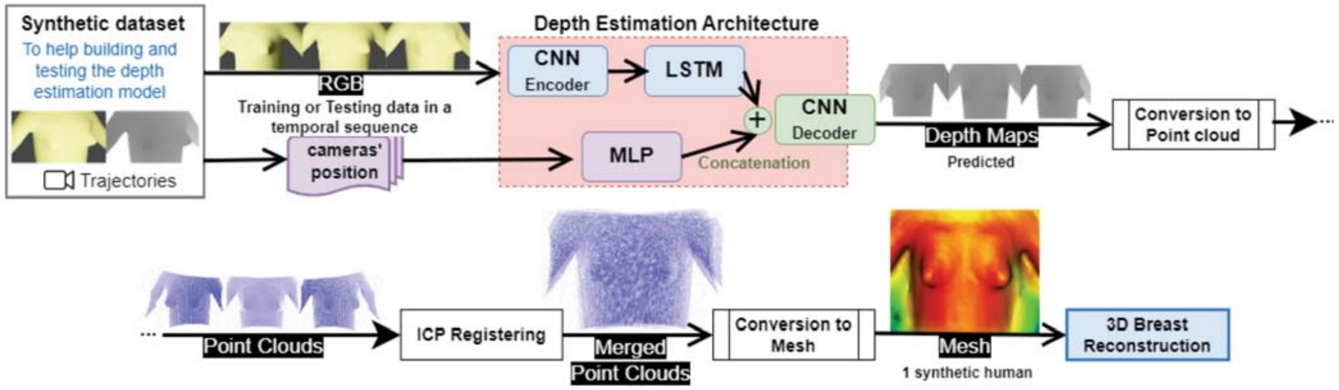

Figure 1 - 3D breast reconstruction pipeline

TABLE I. COMPARISON OF THE RESULTS OBTAINED FOR 3 ARCHITECTURES FOR EACH ONE'S BEST EPOCH, WHERE DM AND CP CORRESPONDS TO DEPTH MAP AND CAMERAS' POSITION, RESPECTIVELY

Architecture	Dataset	d_1	d_2	d3	rel	rms	log10	Mean error (mm)	Min error (mm)	Max error (mm)
CNN (epoch 60)	RGB + DM	0.9905	0.9965	0.9999	0.0168	0.0150	0.0068	4.5	1.2	34.9
CNN + LSTM (epoch 45)	RGB + DM	0.9983	0.9998	1	0.0122	0.0149	0.0053	3.2	1.18	17.9
CNN+LSTM+MLP (epoch 50)	RGB + DM + CP	0.9979	0.9998	1	0.0130	0.0146	0.0056	2.9	1.0	14.0

$$L(gt, pred) = L_{SSIM}\,(gt, pred) + L_{Grad}\,(gt, pred) + \theta L_{L1}\,(gt, pred) \qquad (2)$$

This work used the DenseNet-161 as the CNN encoder since it proved to be the best for depth estimation using this synthetic dataset [12]. The 4-dimensional input tensor aggregates the RGB images information. The output is a group of 13 feature maps. The last feature map, with a size of batch size×2208×15×20, is the LSTM's input. This tensor passes through a maxpool layer, enters the LSTM module, and is upsampled. To include the cameras' position, an MLP with 3 linear layers was added. Its input corresponds to the cameras' position (x,y,z). The outputs from the LSTM and the MLP modules are then concatenated. The CNN's decoder, which is constituted of 2 different convolutional layers and 4 distinct up-sample functions, receives this merged tensor. All the up-sample functions share the same structure, however, their inputs and output differ from each other. The decoder output is a final tensor with size: batch size×1×240×320, which consists of a depth map with a single channel.

B. *From Depth Maps to 3D Reconstruction*

To convert each depth map to a point cloud, this implementation needed to determine the 3D coordinates for each pixel. The depth value, z, can be extracted from the depth map, however, since the depth map codifies the depth in relation to the camera plane and not the optical center axis, to determine the x and y, every pixel had to be converted to its equivalent angle in the lens. This work calculated the constant that allows the conversion between pixels to angles, β, (3) using the horizontal field of view (HFOV).

$$\beta = \frac{\tan\left(\frac{\mathrm{HFOV}}{2}\right)}{\frac{\mathrm{width}}{2}} \qquad (3)$$

For each pixel belonging to the depth map, its value of x and y was converted to quadrants in relation to the center of the image, obtaining x_q and y_q. Then, by applying (4), the corresponding angle for x_q and y_q was determined, which corresponds to the angle of deviation between the correspondent pixel and the center axis.

$$\begin{aligned} angle\ x &= \arctan(x_q \times \beta) \\ angle\ y &= \arctan(y_q \times \beta) \end{aligned} \qquad (4)$$

To finally obtain the point cloud, the 3D coordinate x, pc_x, and the 3D coordinate y, pc_y, were determined with (5). The 3D coordinate z, pc_z, is already contained on the depth map.

$$\begin{aligned} pc_x &= (pc_z \times \tan(angle\ x)) \\ pc_y &= (pc_z \times \tan(angle\ y)) \end{aligned} \qquad (5)$$

To align the different point clouds, the Iterative Closest Point (ICP) algorithm was implemented. This determines the transformation to overlap two point clouds, allowing to merge them after. This procedure was applied to 7 individual point clouds to create a final merged one, for each synthetic human, as depicted in Figure 2. After this, each merged point cloud was imported to *Cloud Compare* software [18], where the Statistical Outlier Removal (SOR) filter was applied, the point cloud's normal were estimated with Hough transform, and the *PoissonRecon* plugin was employed, to generate the meshes.

IV. EXPERIMENTS AND RESULTS

To study the benefits of the proposed strategy, several experiments were performed. The DL architectures were implemented using the Pytorch framework and the ADAM optimizer. It was run on an NVIDIA A100 40GB GPU. The Blender's camera HFOV and the depth map width were set to 0.691 radians and 320, respectively. The employed synthetic dataset has 20 and 10 humans for training and testing, respectively, and each human has 200 images captured at 20 horizontal and 10 vertical positions.

A. *Depth Estimation Architectures*

To study the DL segment additions, this work compared 3 different DL architectures constituted by: the described CNN; the CNN and LSTM; the CNN, LSTM, and MLP. These architectures were trained for 80 epochs, using the synthetic dataset. Then, the best epoch for each one was determined as shown in Table I, where *rel*, *rms*, and *log10*, correspond to the average relative error, the root mean squared error, and the average error using logarithmic of base 10, respectively, and d_1, d_2, and d_3 correspond to the threshold accuracy of values 1.25, 1.25^2 and 1.25^3, respectively.

B. *3D Reconstruction*

The RGB images and depth maps that constitute the synthetic dataset, employed in this work, were acquired around synthetic humans. These humans are in the mesh format and can be used to obtain the entire procedure's error. Therefore, this work generated the ground truth and predicted meshes but the final comparison is done between the predicted meshes and the 10 testing synthetic humans, as represented in Figure 3 and Figure 4, because this corresponds to the entire implementation error. This error is presented in Table II and was obtained using the *Cloud Compare* software. For that, the ICP tool was employed, which added intrinsic error.

V. DISCUSSION

Regarding the DL architecture, it is possible to observe that all the sequential additions generated better results as demonstrated in Table I. The best architecture includes the CNN, LSTM, and MLP, achieving a mean error of 2.9 mm, and the lowest maximum and minimum error. As shown in

TABLE II. MEAN DISTANCE IN MM BETWEEN THE PREDICTED AND SYNTHETIC HUMAN MESHES FOR THE TESTING DATA, USING CLOUDCOMPARE.

Humans Prototypes	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	All data	Without H3
Mean Distance(mm)	2.4	2.2	11.2	3.8	6.3	4.6	4.3	2.9	3.5	5.3	4.6	3.9
Standard deviation (mm)	1.6	1.9	8.9	3.2	5.1	3.9	4.6	2.5	3.8	3.9	2.4	1.2

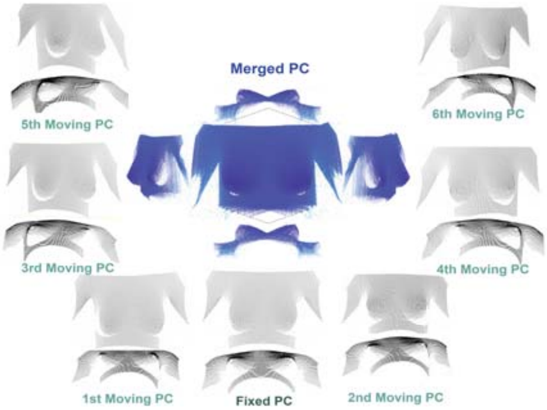

Figure 2 - Example of one merged point cloud, for one of the humans, created from 7 different point clouds, where PC stands for point cloud.

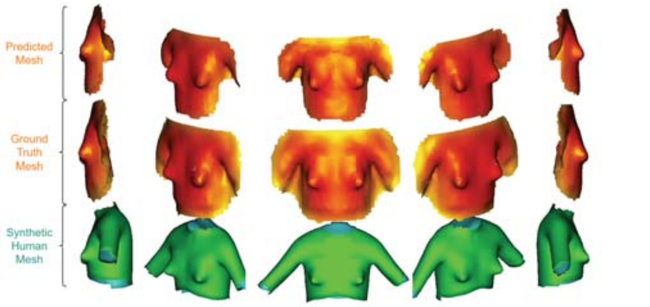

Figure 3 - Comparison between the predicted, ground truth and corresponding synthetic human meshes.

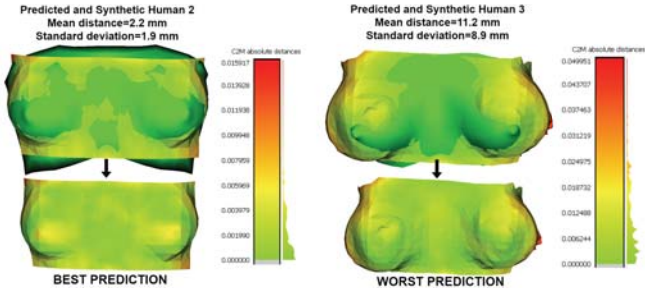

Figure 4 - Comparison between the best and worst predicted mesh overlapped with the respective synthetic human meshes for the testing data.

Table I, the addition of the LSTM reduces the high outliers and mean error present in the CNN architecture by promoting better temporal information. The introduction of an MLP segment allowed the addition of the cameras' position, which improved the spatial information and the predictions.

During the conversion stage, the merging process allowed increasing the information in some breast areas that were not well represented in the 7 individual point clouds, as depicted in Figure 2. In Figure 3, it is observable that the predicted and the ground truth meshes have difficulty in representing the nipple region. Since the ground truth mesh shows problems too, this must be caused during the creation of the merged point clouds. The predicted mesh is also affected by the depth map prediction which offers errors in the nipple area. Nevertheless, the predicted and synthetic human breasts are very similar in terms of size and shape. As represented in Table II, the mean error and standard deviation for the 10 humans correspond to 4.6 mm and 2.4 mm, respectively. But if the third human is excluded since that, in comparison with the others, it shows a very discrepant error, the mean error and the standard deviation are 3.9 mm and 1.2 mm, respectively. The third human has a higher error because, as depicted in Figure 4, breasts that are too pointy, too big, and that extend beyond the thorax, are harder to reconstruct. Contrarily, the smaller breasts with reduced pointiness are the easiest ones. Therefore, it is possible to infer that breast physiognomy affects the amount of error in the predictions and that the distribution and type of error are directly related to the breast zone. This can happen because the dataset does not possess enough large breasts and the merging process of the point clouds requires more individual point clouds. Both issues will be mitigated in the future.

VI. Conclusion

This work proposes a pipeline to obtain a 3D breast reconstruction from monocular RGB images. This can be later used for trajectory planning to enable robotic-guided breast biopsy. This work validated the improvement that arises from combining a CNN, LSTM, and MLP, allowing the addition of the cameras' position and the increase of the temporal and spatial information. Furthermore, it is demonstrated that it is possible to generate meshes from depth maps.

References

[1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/CAAC.21660.

[2] "World Cancer Day 2021: Breast cancer overtakes lung cancer in terms of number of new cancer cases worldwide – IARC."https://www.iarc.who.int/infographics/world-cancer-day-2021/ (accessed Jan. 30, 2023).

[3] "Types of Breast Cancer - National Breast Cancer Foundation." https://www.nationalbreastcancer.org/types-of-breast-cancer/ (accessed Jan. 30, 2023).

[4] "Breast Cancer Facts & Statistics 2023." https://www.breastcancer.org/facts-statistics (accessed Jan. 30, 2023).

[5] "The Radiology Assistant : Bi-RADS for Mammography and Ultrasound 2013." https://radiologyassistant.nl/breast/bi-rads/bi-rads-for-mammography-and-ultrasound-2013 (accessed Jan. 30, 2023).

[6] E. Łukasiewicz, A. Ziemiecka, W. Jakubowski, J. Vojinovic, M. Bogucevska, and K. Dobruch-Sobczak, "Fine-needle versus core-needle biopsy – which one to choose in preoperative assessment of focal lesions in the breasts? Literature review," *J Ultrason*, vol. 17, no. 71, pp. 267–274, Dec. 2017, doi: 10.15557/JOU.2017.0039.

[7] B. Oliveira, P. Morais, H. R. Torres, A. L. Baptista, J. C. Fonseca, and J. L. Vilaça, "Characterization of the Workspace and Limits of Operation of Laser Treatments for Vascular Lesions of the Lower Limbs," *Sensors*, vol. 22, no. 19, p. 7481, Oct. 2022

[8] N. Costa *et al.*, "Augmented Reality-Assisted Ultrasound Breast Biopsy," *Sensors*, vol. 23, no. 4, p. 1838, Feb. 2023, doi: 10.3390/s23041838.

[9] V. Groenhuis, F. J. Siepel, J. Veltman, J. K. van Zandwijk, and S. Stramigioli, "Stormram 4: An MR Safe Robotic System for Breast Biopsy," *Ann Biomed Eng*, vol. 46, no. 10, pp. 1686–1696, Oct. 2018, doi: 10.1007/S10439-018-2051-5/TABLES/3.

[10] D. Navarro-Alarcon *et al.*, "Developing a Compact Robotic Needle Driver for MRI-Guided Breast Biopsy in Tight Environments," *IEEE Robot Autom Lett*

[11] M. K. Welleweerd, F. J. Siepel, V. Groenhuis, J. Veltman, and S. Stramigioli, "Design of an end-effector for robot-assisted ultrasound-guided breast biopsies," *Int J Comput Assist Radiol Surg*, vol. 15, no. 4, pp. 681–690, Apr. 2020

[12] B. Duarte, B. Oliveira, H. Torres, P. Morais, J. Fonseca, and J. Vilaça, "Augmented Synthetic Dataset with Structured Light to Develop Ai-Based Methods for Breast Depth Estimation", doi: 10.1145/3569192.3569206.

[13] J. L. Schonberger and J. M. Frahm, "Structure-from-Motion Revisited," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 4104–4113, Dec. 2016.

[14] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y. G. Jiang, "Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images," *Lecture Notes in Computer Science* vol. 11215 LNCS, pp. 55–71, Apr. 2018, doi: 10.48550/arxiv.1804.01654.

[15] W. Bian, Z. Wang, K. Li, and V. A. Prisacariu, "Ray-ONet: Efficient 3D Reconstruction From A Single RGB Image," Jul. 2021,

[16] M. M. Ahsan, T. E. Alam, T. Trafalis, and P. Huebner, "Deep MLP-CNN Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients," *Symmetry 2020, Vol. 12, Page 1526*, vol. 12, no. 9, p. 1526, Sep. 2020.

[17] I. A. Kaust and P. Wonka, "High Quality Monocular Depth Estimation via Transfer Learning", Accessed: Jan. 31, 2023. [Online].

[18] "CloudCompare (GPL) · GitHub." https://github.com/CloudCompare (accessed Jan. 31, 2023).