Oliver Gölz*, Julia Chen, Clemens Schlegel, Jens Langejürgen

SonoMap – Automated Ultrasound Documentation in Paediatric Care

https://doi.org/10.1515/cdbme-2025-0162

Abstract: Pictograms are commonly used to document probe positions in diagnostical ultrasound procedures. A new computer-vision based method using depth cameras is proposed and tested to automatically document the 3D probe position relative to the patient body. The probe is held in different poses by the examiner and translation magnitudes and tilting angles are compared to reference values obtained by a robotic arm. Our results express the systems general suitability for visual documentation. Further optimization of the processing pipeline for this use case is planned, followed by a study in a clinical environment.

Keywords: ultrasound, documentation, pediatric care, depth camera, digitalization, computer vision, pictograms

1 Introduction

Ultrasound (US) is a widely utilized imaging modality, particularly favoured over CT-scans in paediatric care, as it does not expose patients to harmful radiation [1]. One drawback is the lack of spatial information regarding the position of the US image relative to the patient body. To date, pictograms are utilized to document the spatial position of the US probe relative to the patient body (see Figure 1), e.g. to areas of interest such as space-occupying lesions. The patient body is often simplified by landmarks given in the relevant context. To annotate this pictogram, manual input is required. In most devices the practitioner uses a trackball to enter the position of the probe relative to the patient body. This requires time consuming and possibly inaccurate manual visual matching of the hand position with the 2D pictogram. Additionally, spatial information is lacking. While the 2D pictogram roughly captures the (x,y) position of the probe, no tilting angles are preserved. These could be beneficial for future visits involving the same case.

Figure 1: Abdominal pictogram used in current practice

Different solutions have been explored to more accurately document the position of the ultrasound probe. Jiang et al. [2] present a solution in the domain of breast cancer documentation, which uses an electromagnetic tracking system, including a transponder attached to the ultrasound probe. The 3D position is then referenced against anatomical landmarks and mapped to a 3D visualisation or a 2D pictogram. While being accurate the method requires expansive hardware and anatomical landmarks to document the scene. Tracking options such as optical tracking with passive or active trackers or with inertial sensors are possible too [3]. The drawback is that these do not capture the patients' surface and require reference points. Another idea proposed by Fröhlich et al. [4] includes attaching a miniature camera to the probe, which captures the surrounding scene. Their study showed that images from the camera improve the quality of the documentation and saves time during examinations. This solution is affordable, but the spatial documentation is lacking, as the camera only acquires closeup images of the probe's surroundings without the full 3D pose.

Depth cameras using technologies such as LIDAR and stereoscopy are commonly applied in many domains where accurate 3D mapping is required. To date they have not been used in clinical practise to improve documentation in the field of ultrasound care. For this reason, we propose a computer vision-based method to document the spatial position of the ultrasound probe utilizing depth cameras and compare it to robotic reference values. We present the data processing pipeline combining 2D image segmentation and 3D point cloud manipulation using a stereoscopic camera with the goal of generating a 3D visualisation of the patient specific probe position. Future use cases, first test results and potential shortcomings are discussed in this paper.

***Oliver Gölz** Fraunhofer IPA, Mannheim, Germany, oliver.goelz@ipa.fraunhofer.de
Julia Chen, Clemens Schlegel, Jens Langejürgen, Fraunhofer IPA, Mannheim, Germany

2 Methods

2.1 Measurement Setup

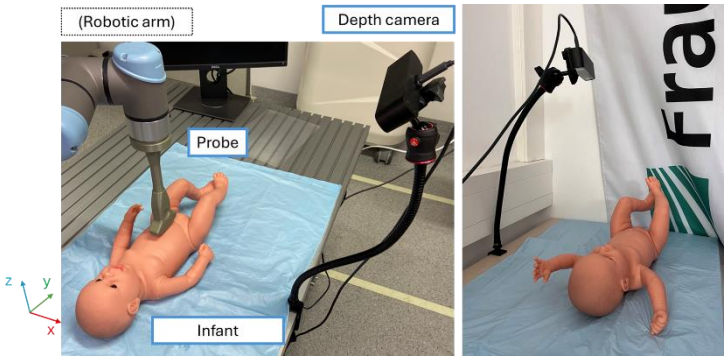

Figure 2: Left - Measurement setup with stereoscopic camera and added robotic arm for precise reference movements, right - Intended paediatric care use case

The main component of our system is the commercially available ZED 2i stereoscopic depth camera (Stereolabs), which is attached to the table with a full field of view of an infant phantom and a 3D printed true to size US probe (c60e). The ZED 2 is set to 1080p, 15 fps with depth mode ULTRA. For our tests the probe mockup is attached to the flange of a 6 DoF robotic arm UR5e (Universal Robots). The robot is not an integral part of the system; rather, it is utilized to provide accurate reference poses of the ultrasound probe.

2.2 Data Processing Pipeline

The main task of the processing pipeline is the generation of a patient-specific 3D visualisation using the US probe's pose during image acquisition. The captured scene includes the environment as well as the examiners arms and hands. To simplify the scene to only contain the patient and ultrasound probe, point cloud segmentation is required.
We segment the objects of interest in the 2D RGB image and project the masks to the 3D depth matrix captured in the same camera pose. The depth matrices are converted to point clouds. The pose of the US-probe is obtained and used to register a ground truth 3D object of the utilized probe into the scene. Each step as shown in Figure 3 is explained in detail.

2D image segmentation
We deploy a pretrained YOLO 11n-seg model (Ultralytics) [5] which is further trained for our use case using transfer learning. Segmentation mask for the classes "infant", "probe" and "examiner" are semi-automatically annotated using

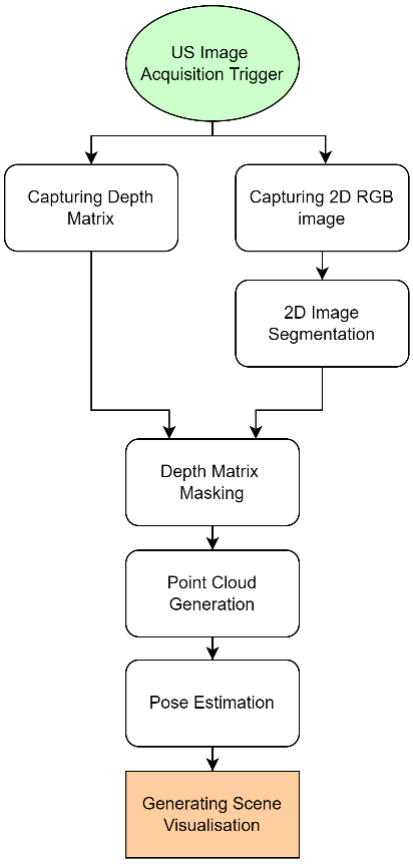

Figure 3: RGB and depth image processing pipeline

SAM2 [6]. The examiner mask includes hand and arm of the practitioner holding the probe. A dataset of 120 images of the scene as seen in Figure 2 is used initially for the transfer and trained over 300 epochs with the default YOLO 11 training parameters. The infants and cameras pose and position on the table remain constant, while the probe pose is varied under realistic conditions. The added shaft of the c60e probe mockup is not segmented in our training data.

Depth matrix segmentation
Both 2D RGB images and the depth matrix are acquired simultaneously with the same camera pose. This enables masking of the depth matrix via the predicted 2D segmentation masks. For each object of interest, only the depth values corresponding to the masked areas are preserved. The point clouds P_{infant} and P_{probe} are generated from each masked depth matrix while colours are preserved. Any extreme outliers are removed afterwards.

Pose estimation of ultrasound probe
At this stage, P_{infant} and P_{probe} are expressed in the camera coordinate system CS_{camera}.

Using principal component analysis (PCA), P_{probe} is translated so that its tip lies in the origin of CS_{camera} and a rotation matrix is obtained which orientates the point cloud along its principal components. The inverse of this resulting transformation matrix $T_{camera,probe}$ can register an imported ground truth probe model $P_{probemodel}$ given in CS_{camera} to the pose of P_{probe}:

$$P_{probe} = P_{probemodel} \cdot T_{camera,probe} \quad (1)$$

2.3 Measurement Procedure

In a real-world setup, the pose of the probe relative to the patient point cloud is most relevant for documentation purposes. In this first measurement setup we use the tooltip pose $T_{robotbase,TCP}$ of the probe mockup expressed in $CS_{robotbase}$ as a ground truth reference to $T_{camera,probe}$ obtained by our system. To compare the poses $T_{camera,probe}$ needs to be transformed to $CS_{robotbase}$, which requires the following transformations:

$$T_{robotbase,probe} = T_{robotbase,world} \cdot T_{world,camera} \cdot T_{camera,probe} \quad (2)$$

$T_{world,camera}$ is a rotation matrix obtained by the ZEDs onboard IMUs and describes the cameras pose. The accuracy of the cameras pose estimation is not publicly available. $T_{robotbase,world}$ is simplified by a rotation matrix without a translation component as the camera is in unspecified distance to the $CS_{robotbase}$ origin.

Initially, ten estimations of $T_{camera,probe}$ are obtained with $T_{robotbase,TCP}$ aligned with the z-axis. Afterwards, the probe is manually moved to various poses and 10 positions are documented without a visible examiner. Ten additional positions are documented with the probe held and partially obstructed by the examiner. The magnitude of the translation vector $\vec{s}$ is calculated by subtracting the mean starting position. The deviations between the vector magnitudes of $T_{robotbase,probe}$ and $T_{robotbase,TCP}$ are calculated.
Additionally, the probe is robotically tilted along the $CS_{robotbase}$ x-axis and y-axis ten times, respectively, by varying magnitudes between ± 40°. The angle between the main axis of $P_{probemodel}$ and the z-axis of $CS_{robotbase}$ is documented. The baseline deviation for 0° tilt is measured over 10 pipeline iterations.

3 Results

Representative quantitative results of three poses including one with a visible examiner are visualised in Figure 4. The translation vector magnitude deviations between our system and the robots ground truth are shown in Table 1. The deviations between our systems tilting angle (z-axis to probe main axis) and the robots tilt are listed in Table 2.

Segmentation: Both the infant and probe can be segmented with high confidence (> 0.95) in our validation dataset (n=10). Over- or Undersegmentation is observable around the object's edges. The shaft of the mockup probe is not part of the mask.

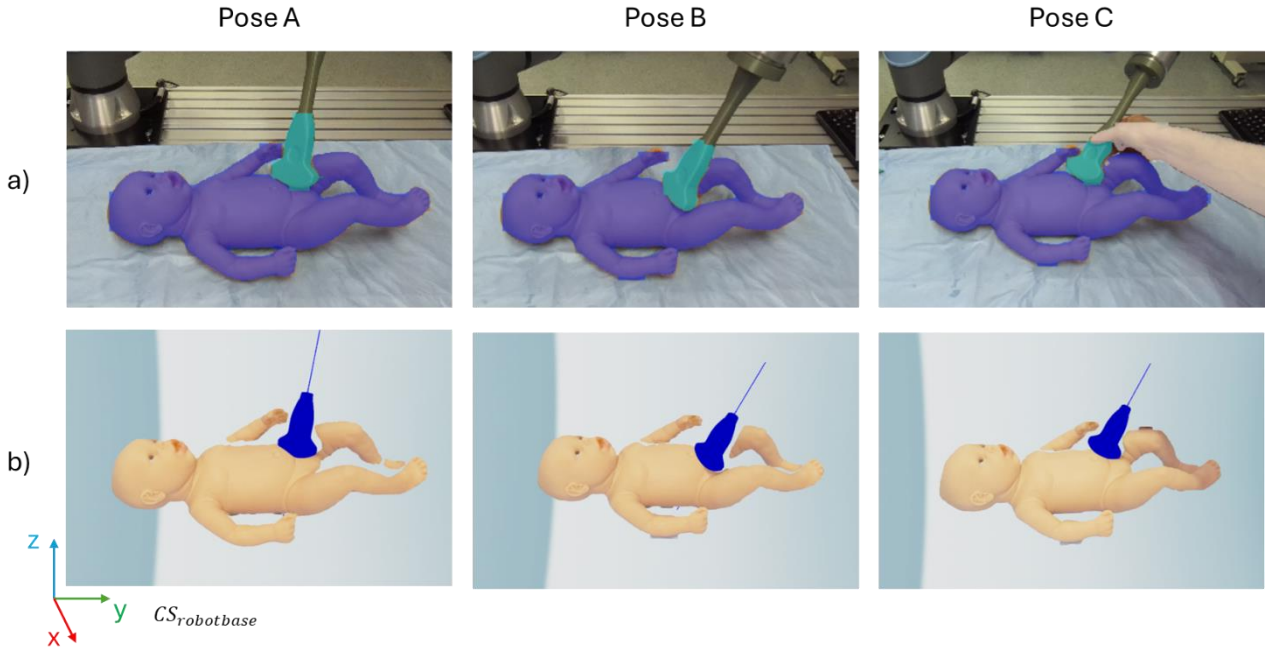

Figure 4: Comparison of different poses with a) 2D RGB Segmentation Masks and b) 3D Visualisation of $P_{probemodel}$ and P_{infant}

Pose Estimation: The (x,y,z) euler angles of $T_{world,camera}$ obtained from the cameras IMUs are: -49.80 ± 0.13°, 0.00 ± 0.00°, -0.51 ± 0.05° (n=10). The mean position of the tip of $P_{probemodel}$ expressed in $CS_{robotbase}$ is: -514 ± 0.9 mm, 79 ± 1.6 mm, -548 ± 0.6 mm (n=10).

Table 1: Magnitude deviations of $\vec{s}$ with n=10 per column

Probe to Robot $\Delta\vec{v}$	No examiner $\Delta\vec{s}$ (mm)	With examiner $\Delta\vec{s}$ (mm)
Mean	12.7	13.7
STD	6	7.4
Max	24.8	26.3

Table 2: Probe tilt deviations with n=10 per column

Probe main axis to z-axis	Robot No tilt (°)	Robot x-axis tilt (°)	Robot y-axis tilt (°)
Mean	3.2	2.4	2.9
Max	4	6	9
STD	0.8	2	2.6

4 Discussion

Both the translation vector magnitude deviations and the tilt deviations seem generally acceptable for a purely visual documentation with manual probe positioning. The obstruction by the examiner results in a slightly increased but tolerable translation deviation. It is not possible to compare the complete 6 DoF probe pose with this setup. A separate comparison of x-axis and y-axis tilt deviations in our systems probe pose is currently missing. IMU and camera calibration might cause additional deviations in $T_{robotbase,probe}$ although the visible probe pose in the camera coordinate system seems to be quite accurate. To improve segmentation, additional training data with finer annotations and more variations in the scenes surroundings and illumination should be generated.

In such a paediatric care application, it might not be possible to store a 3D depth image of a patient without any anonymisation. To solve this a generic patient model could be fitted to the point cloud or other methods of patient anonymisations can be explored. The effects of the infants' movements on the estimation are also currently unclear.

From an application perspective, our system could be easily integrated with existing ultrasound solutions, offering a cost-effective approach to document ultrasound examinations. Use of this system may lead to more efficient follow-ups and more reproducible diagnoses. To evaluate its effectiveness, a user study is planned to compare the time required to locate the probe position using pictograms versus 3D visualization.

5 Conclusion

Our proposed automated documentation system may render the manual generation of pictograms obsolete. A 3D documentation would provide more spatial information while enabling the generation of a patient-individualized visualisation. The results express a general suitability of this system for visual documentation. Further optimization of the processing pipeline is planned and followed by a first study in a clinical environment.

Author Statement

Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] P. F. Hoyer, I. Finkelberg, B. Prusinskas, and M. Cetiner, "Update Ultraschall in der Kinder- und Jugendmedizin," *Monatsschr Kinderheilkd*, vol. 172, no. 12, pp. 1096–1110, 2024, doi: 10.1007/s00112-024-02082-9.

[2] W.-W. Jiang, C. Li, A.-H. Li, and Y.-P. Zheng, "Clinical Evaluation of a 3-D Automatic Annotation Method for Breast Ultrasound Imaging," *Ultrasound in medicine & biology*, vol. 42, no. 4, pp. 870–881, 2016, doi: 10.1016/j.ultrasmedbio.2015.11.028.

[3] C. Peng, Q. Cai, M. Chen, and X. Jiang, "Recent Advances in Tracking Devices for Biomedical Ultrasound Imaging Applications," *Micromachines*, vol. 13, no. 11, 2022, doi: 10.3390/mi13111855.

[4] E. Fröhlich *et al.,* "Piktocam statt Piktogramm - Validierungsstudie zur Qualitätsverbesserung der Bilddokumentation beim abdominellen Ultraschall," (in ger), *Zeitschrift fur medizinische Physik*, vol. 26, no. 3, pp. 251–258, 2016, doi: 10.1016/j.zemedi.2016.01.001.

[5] Glenn Jocher and Jing Qiu, *Ultralytics YOLO11*, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

[6] N. Ravi *et al.,* "SAM 2: Segment Anything in Images and Videos," 2024. [Online]. Available: https://arxiv.org/abs/2408.00714