TU WIEN Informatics

Student Project

Stereoscopic Volume Rendering of Volumetric Cardiac Ultrasound with the Sony Spatial Reality Display

Sebastian Antes

Supervisor:

Dr.techn. Alexey Karimov

Siemens Healthineers

Faculty of Informatics

TU Wien

Vienna, April 30, 2026

Sebastian Antes | Dr.techn. Alexey Karimov

Erklärung zur Verfassung der Arbeit

Sebastian Antes, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel" habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 30. April 2026

Sebastian Antes

Abstract

While 3D cardiac ultrasound is essential for modern diagnostics, viewing volumetric data on 2D screens forces clinicians to perform complex mental reconstructions of anatomy. This project presents a volume rendering pipeline for the Sony Spatial Reality Display (SRD) [Son25], providing a glasses-free, stereoscopic visualization. The primary objective is to facilitate the spatial assessment of complex cardiac anatomy by enabling precise, real-time distance and size measurements within a stereoscopic volumetric environment.

The technical solution utilizes a custom Unity-based raymarching shader Beer-Lambert accumulation [Max95] to visualize intensity fields. To overcome Unity Render Graph limitations, a specialized `SRDWriteBackFeature` was developed using an unsafe rendering pass to bind a GPU buffer to a Unordered Access View (UAV) slot. This enables a shader-integrated picking system that calculates Euclidean distances in real-time between markers placed by the user on the volumetric surface.

Although hardware sensor failure prevented a quantitative user study, technical validation confirms a stable rendering pipeline capable of real-time clinical interaction. Qualitative results indicate that the "holographic" rendering significantly improves spatial perception compared to traditional 2D monitors. Future developments should focus on sterile interaction methods, such as foot pedals, to facilitate adoption in interventional environments.

Contents

CHAPTER 1

Introduction

In modern cardiology, the transition from 2D to 3D and 4D ultrasound has significantly improved how we capture the moving heart. However, a major bottleneck remains: these volumetric datasets are still primarily viewed on conventional 2D monitors. This requires clinicians to mentally reconstruct three-dimensional structures from flat slices, which is both cognitively demanding and prone to error. This project explores an alternative approach by using the Sony Spatial Reality Display (SRD) [Son25] to visualize cardiac ultrasound data in a glasses-free, stereoscopic 3D environment.

In interventional cardiology, the ability to accurately measure internal diameters and distances within the heart is vital for procedural planning. Currently, clinicians often rely on 2D slices to estimate 3D dimensions. This system aims to improve this workflow by providing a 'holographic' interface where users can interact with the volume and perform measurements naturally, leveraging depth cues to improve spatial understanding.

The goal of this work was to develop a functional rendering pipeline in Unity that integrates the Sony SRD with volumetric ultrasound data. The implementation focuses on a custom raymarching shader to handle the volume data and the synchronization of the display's eye-tracking for a stable 3D effect. By incorporating a 3D marker system, this project demonstrates how autostereoscopic displays can enhance spatial perception and measurement accuracy in clinical cardiac workflows.

CHAPTER 2

Related Work

Understanding medical volumetric data using 2D displays can be really challenging, as clinicians have to mentally reconstruct 3D anatomy. To address these limitations, there is a growing interest in alternative visualization technologies that provide more natural spatial perception. This chapter explores methods for visualizing volumetric data using various emerging display technologies beyond conventional 2D screens, including virtual and mixed reality (VR/MR) as well as glasses-free 3D displays such as the Sony Spatial Reality Display [Son25] used for this project. The focus here is on medical applications, particularly the visualization of cardiac ultrasound data.

Wheeler et al. proposed a Unity-based VR measurement tool specifically implemented for 3D echocardiography. Linear measurements were shown to be statistically equivalent to a reference 2D platform while outperforming a commercial 3D tool, indicating that immersion can preserve quantitative reliability when VR is used [WDP+19].

Kim et al. focused on enabling group discussions about congenital heart disease (CHD) by developing a multiuser VR system. In their study, groups of 4-5 people discussed diagnostic cases of CHD of varying structural complexity using conventional 2D displays and the group VR software. The highest diagnostic accuracy and user preference was achieved when using the full-immersive VR system, particularly for complex anatomy [KLM+20]. Mixed-reality (MR) approaches extend VR by anchoring volumetric data in the physical environment. By rendering 4D (3D + time) cardiac ultrasound as holograms, Maddali et al. found that MR improved spatial orientation and morphological understanding compared to a standard clinical viewer, although the image quality was perceived lower [MBK+22].

The head-mounted devices we just discussed come with ergonomic and sterility contraints, motivating glasses-free alternatives for clinical usecases.

Kang et al. proposed a glasses-free 3D autostereoscopic display system which uses an eye tracking algorithm based on machine learning techniques to display 3D images. When rendering volumetric data of coronary computer tomography (CT), expert readers were able to identify arterial structures faster than with conventional 2D displays without reporting discomfort [KCH22]. Aoki et al. introduced the Retinal Homing Display, a head-tracked autostereoscopic retinal-projection system that provides correct stereoscopic alignment without the need of the user wearing glasses [AHIR23]. Both approaches use the positions of the head and eyes to render the views in real time. As the viewpoints are conditioned on the tracked eye position, the spatial perception on multi-view hardware is maintained correctly [KCH22, AHIR23]. Similarly, Sony´s spatial reality displays combine stereo generation with viewer tracking to provide a holographic rendering of the spatial data [Son25].

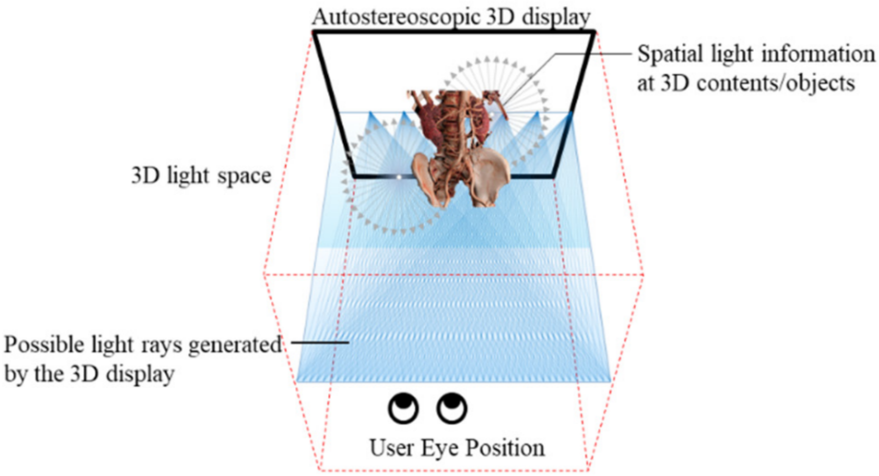

Figure 2.1: Eye tracking-based light-field 3D display concept by Kang et al. [KCH22], showing the generation and modeling of directional light rays.

CHAPTER 3

Overview of the Approach

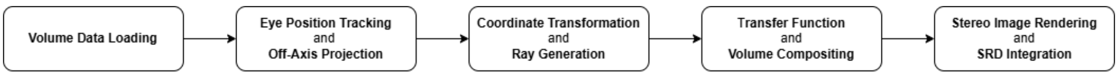

Figure 3.1: Pipeline Overview

The visualization pipeline used consists of five main stages: data loading, eye position tracking, coordinate transformation and ray generation, compositing and finally rendering.

Volume Data Loading. Cardiac ultrasound data are provided as volumetric intensity fields in MetaImage format (.mhd/.raw). The VolumeLoader component parses the metadata to extract voxel dimensions, spacing and data type, then constructs a 3D texture that encodes the ultrasound volume. This texture serves as the input to a custom raymarching shader.

Eye Position Tracking and Off-Axis Projection. The Sony SRD [Son25] continuously tracks the viewer´s head and eye positions relative to the display surface. These data are used to compute off-axis projection matrices for each eye, ensuring geometrically correct stereo parallax as the viewer moves. The two virtual cameras (left and right) inside Unity are updated accordingly, maintaining alignment between virtual and physical viewing frustums.

Coordinate Transformation and Ray Generation To ensure the volumetric data aligns with the physical display, the system translates the viewer's eye positions into the virtual object space of the ultrasound volume. The `SRDVolumeManager` resolves the world-space positions of the left and right eye cameras provided by the Sony sdk. These positions are transformed into the local object space of the volume renderer using Unity's `unity_WorldToObject` matrix.

For each fragment, the system first maps the screen-space UV coordinates to Normalized Device Coordinates (**NDC**), which represent the viewing volume in a range from -1

to 1 across all axes. These NDC values are then transformed by the inverse of the SRD's per-eye projection matrix to obtain a view-space direction. Finally, this vector is transformed into world space and subsequently into the local object space using the inverse view and `unity_WorldToObject` matrices.

The system supports two distinct ray generation paths to handle different viewing requirements:

- **SRD Stereo Mode**: This is the primary mode where unique ray directions are calculated for each eye using the tracked eye positions.

  It utilizes the `unity_StereoEyeIndex` to fetch the correct per-eye inverse projection and view matrices every frame.

- **Mono from SRD Mode**: A fallback or debugging mode that forces the shader to use the left-eye data (eye index 0) for both rendering passes. This effectively removes the horizontal parallax, providing a stable, non-stereoscopic image within the SRD environment.

In the fragment shader, a classic "slabs test" determines the intersection points (t_{enter} and t_{exit}) where the ray enters and leaves the unit cube containing the volume data.

Transfer Function and Volume Compositing The visualization of internal cardiac structures depends on a custom raymarching loop that accumulates color and opacity. The system uses a **Transfer Function (TF)**, managed by the `SRDVolumeManager`, which maps raw ultrasound intensity values to specific RGBA values. In this model, the Alpha channel of the TF is treated as an extinction coefficient (σ), while the RGB channels represent the emissive color. To optimize performance and image quality, the shader implements two key techniques:

- **Adaptive Step Sizing**: The raymarcher takes larger steps in low-opacity regions to save performance, while slowing down in dense areas to capture detail.

- **Beer-Lambert Accumulation**: Light absorption and emission are calculated using an exponential decay function to simulate realistic tissue density [Max95]. The accumulated color C_{acc} and opacity A_{acc} are updated at each step i with length Δt using the extinction coefficient σ derived from the Transfer Function:

$$\alpha_i = 1 - e^{-\sigma \cdot \Delta t} \tag{3.1}$$

$$C_{acc} = C_{acc} + (1 - A_{acc}) \cdot C_i \cdot \alpha_i \tag{3.2}$$

  This provides a realistic "glow" effect where internal cardiac structures are attenuated by the surrounding density.

Stereo Image Rendering and SRD Integration The integration with the Sony SRD requires a specialized rendering path to prevent visual artifacts. Because the SRD uses a single-pass instanced stereo approach, both eyes share the same horizontal render target [Uni25]. To prevent the right eye from accidentally sampling the left eye's view, a **Stereo UV Adjustment** is implemented.

Using Unity's `UnityStereoTransform-ScreenSpaceTex` macro, the texture coordinates are re-centered based on the active eye index. The `SRDVolumeManager` acts as the bridge, injecting the correct view and projection matrices into the shader every frame.

Measurement Marker Placement To support quantitative analysis of the volumetric data, the system includes a 3D picking system. This is achieved through a **GPU Writeback** technique:

1. **Cursor Tracking**: The `CuttingPlane-Measurement-Controller` calculates the mouse position in 3D space by applying a homography transformation to the screen UVs.
2. **Depth Picking**: When the user hovers over the volume, the fragment shader identifies the first point where the accumulated opacity exceeds a `PickThreshold`.
3. **Buffer Readback**: This 3D position is written to a `RWStructuredBuffer` on the GPU. The CPU reads this data back using an asynchronous request, allowing the user to place markers and calculate distances with millimeter precision in a 3D environment.

CHAPTER 4

Implementation Details

The technical implementation of this project focused on creating a performant volume rendering pipeline while overcoming specific limitations of the Unity Universal Render Pipeline (URP) and the Sony SRD SDK.

4.1 Volume Rendering and Raymarching

The core of the visualization is a custom raymarching shader that processes a 3D texture. The shader uses a unit cube gameobject as a proxy geometry, which holds all the scripts as components within Unity. The `SRDVolumeManager` component handles the high-level setup, such as generating the transfer function texture from a Unity Gradient and passing optical properties like absorption and emission boost to the GPU. One specific challenge was the `adaptiveStepSize`, which was implemented to ensure that the raymarcher maintains a high frame rate by increasing step length in empty regions of the ultrasound volume.

4.1.1 Local Object Space Raymarching

To ensure the rendering pipeline remains robust regardless of the volume's transformation within the Unity scene, the raymarching process is performed entirely in **Local Object Space**.

Using the `unity_WorldToObject` matrix, the eye positions tracked by the Sony SRD are transformed into the local coordinates of the unit cube proxy geometry. This approach offers two distinct advantages:

- **Sampling Efficiency**: The volumetric data is stored in a 3D texture where coordinates range from 0 to 1. By defining the proxy cube from -0.5 to 0.5 in object space, a simple mapping function $P_{tex} = P_{obj} + 0.5$ allows for direct texture sampling without complex scaling.

- **Spatial Invariance**: The clinician can rotate, scale, or translate the volume to inspect different cardiac angles without requiring the shader to re-calculate world-space bounding boxes or ray directions.

The intersection of the ray with this local volume is calculated using a classic slabs test to determine the entry (t_{enter}) and exit (t_{exit}) points:

$$t_{min} = \frac{B_{min} - R_o}{R_d}, \quad t_{max} = \frac{B_{max} - R_o}{R_d} \tag{4.1}$$

where R_o and R_d are the ray origin and direction in object space, and B represents the axis-aligned bounding box of the unit cube.

4.2 Interactive Cutting Plane

To provide a detailed exploration of internal cardiac structures that may be occluded in a full volumetric view, an interactive cutting plane was implemented.

The `CuttingPlaneController` allows the user to define a planar boundary via an object-space position and rotation, which is then passed to the shader as a plane point $\vec{p}$ and normal $\vec{n}$. During the raymarching process, the shader evaluates the intersection of each ray with this plane using the standard ray-plane equation . The valid raymarching segment $[t_{enter}, t_{exit}]$ is dynamically clipped based on whether the ray origin lies on the user-selected "positive" or "negative" side of the plane . This enables the clinician to "slice" through the cardiac ultrasound volume in real-time to clearly reveal internal cardiac structures and provide a clear surface for subsequent measurement marker placement.

4.3 Bypassing Render Graph Limitations

A significant challenge occurred when implementing the 3D picking system. In modern Unity URP, the Render Graph system manages resources strictly, which makes it difficult to use `SetRandomWriteTarget` for writing data from a fragment shader into a `ComputeBuffer`. To solve this, we implemented the `SRDWriteBackFeature`. This script uses a `ScriptableRenderPass` [Uni23] that creates an "Unsafe Pass" within the Render Graph . By using an `UnsafeGraphContext`, the system can manually bind a `GraphicsBuffer` to a specific Unordered Access View (UAV) slot (register `u1`). This allows the shader to write the object-space position of the detected heart wall directly into the buffer when the ray's accumulated alpha reaches the `_PickThreshold`.

4.4 Cursor Homography and 3D Interaction

Since the Sony SRD is a physical display with specific dimensions, standard mouse input does not automatically align with the 3D "holographic" view.

In the `CuttingPlane-Measurement-Controller`, we implemented a homography transformation. This logic takes the 2D screen coordinates and calculates the correct position relative to the display's physical edges (`LeftUp`, `RightBottom`, etc.) . This ensures that the 3D crosshair precisely follows the user's intended focus point, allowing for the accurate placement of markers.

4.5 Technical Challenges and Limitations

A critical challenge in implementing real-time GPU-based measurements on the Sony SRD is the phenomenon of **Stereo Contention**. Because the display utilizes a single-pass instanced rendering approach, the fragment shader is executed simultaneously for both the left and right eye views to generate the stereoscopic image.

Without intervention, both eye-rays would attempt to write their detected surface hit positions to the same index in the `RWStructuredBuffer` (UAV slot `u1`). This results in a race condition where the CPU might receive fluctuating values from two different perspectives, leading to jittery or incorrect 3D measurement markers.

To resolve this, the implementation utilizes the `unity_StereoEyeIndex` to restrict write-back operations. In the shader, only the eye matching the `_WritebackEye` property (typically the left eye) is permitted to modify the GPU buffer.

This ensures that while both eyes contribute to the visual "holographic" effect, only one consistent geometric perspective dictates the 3D measurement logic.

CHAPTER 5

Validation and Results

This chapter describes the experimental setup designed to evaluate the accuracy and usability of the implemented renderer. Due to a hardware failure of the Sony Spatial Reality Display, which prevented reliable gaze tracking, the full user study could not be conducted. Therefore, this section outlines the intended methodology and provides a qualitative analysis based on preliminary testing and system performance.

5.1 Experimental Design and Concept

The primary objective of the study was to compare the measurement accuracy and mental workload between a conventional 2D monitor and the Sony SRD. The experiment was designed for five participants to perform a series of anatomical measurements on 3D cardiac ultrasound volumes.

- **Task:** Participants are asked to identify and measure the distance between three to four specific anatomical features, such as chamber diameters or vessel widths.
- **Interaction:** To locate these structures, users interact with the system by moving and rotating a virtual cutting plane through the volume. Once the target structure is visible, a 3D cursor is used to place markers directly on the tissue surface. The Transfer Function is pre-optimized for the dataset, and the `_PickThreshold` is set to 0.01 by default to detect the first hit on the visible volume on the cutting plane. Both, Transfer Function and `_PickThreshold` can also be adjusted by the user to enable a customized selection if needed.
- **Measurement:** The system automatically calculates the **Euclidean distance** between pairs of these markers in object space to provide the final measurement and display it in the User Interface.

- **Configuration:** Measurements are to be performed in both a 2D mono mode and the 3D stereoscopic mode. During the study, the supervising person takes the time the participants need for each measurement assignment in both modes.
- **Quantitative Assessment:** The measurement´s time and accuracy of all target features can be compared between the two visualization modes. The system is configured to record the average frame rate (FPS) and log it upon application closure to measure the performance of the renderer.
- **Qualitative Assessment:** After the tasks, users would complete a System Usability Scale (SUS) and a TLX questionnaire to assess the perceived mental load.

5.2 Hypothesis and Preliminary Observations

The central hypothesis of this project is that improved depth perception through autostereoscopic rendering increases the accuracy and speed of 3D measurements. This is particularly relevant for interventional planning, where the precise spatial relationship between cardiac walls dictates the success of a procedure. Preliminary tests during the development phase indicated that the 3D mode allowed for more intuitive marker placement compared to 2D projections. The "holographic" effect helps the user perceive exactly where the cutting plane was positioned relative to the heart walls. However, the hardware issues with the tracking sensor led to significant "jitter" in the 3D viewpoint, which made a fair quantitative comparison between 2D and 3D impossible at this stage.

5.3 System Performance and Qualitative Results

Despite the lack of user study data, the technical implementation proved to be efficient. The raymarching shader maintained a stable frame rate, which is necessary for real-time clinical interaction . The `SRDWriteBackFeature` successfully bypassed the Render Graph limitations to provide real-time distance calculations. The tests during development suggest that the system offers a superior spatial understanding of the anatomy of the human heart compared to traditional screen-based 3D reconstructions.

CHAPTER 6

Conclusion and Future Work

This project successfully implemented a specialized volume rendering pipeline for cardiac ultrasound data on the Sony Spatial Reality Display. By leveraging a custom raymarching shader and a dedicated GPU write-back feature, the system provides a glasses-free, stereoscopic environment for anatomical visualization and measurement.

6.1 Summary of Contributions

The primary contribution of this work is the development of a real-time 3D autostereoscopic volume renderer tailored for cardiological procedures. We demonstrated that:

- A performant volume renderer can be integrated into the Unity Render Graph to provide high-fidelity ultrasound visualization.
- Spatial perception of complex structures is significantly enhanced by autostereoscopic displays compared to traditional 2D screens.
- Precise 3D measurements are achieved through a Shader-integrated picking system that identifies surface hits during the rendering pass and writes them to a `RWStructuredBuffer`.
- The system successfully calculates Euclidean distances directly in object space based on these GPU-derived coordinates.

While the intended user study could not be finalized due to hardware tracking issues, the qualitative results and technical validation show that such a system is highly feasible for clinical integration. The project highlights the potential of "holographic" displays to reduce the cognitive load of mental 3D reconstruction during surgical planning.

6.2 Future Work and Clinical Adoption

Several avenues for future development have been identified to move this prototype toward clinical adoption:

- **Hands-Free Interaction:** In a sterile operating room, manual mouse clicking is impractical. Future iterations should replace mouse input with pedal actions or voice commands to maintain sterility .
- **Gaze-Based Raymarching:** The current system relies on manual cursor placement. Improving the raymarching logic to trace the user's actual gaze position on the image would allow for even more natural "look-at" interaction and automated focus adjustment .
- **Complex Geometry Measurements:** While Euclidean distance provides a basic diameter, clinical workflows often require measuring the circumference or area of irregular anatomical structures. Extending the marker system to support spline-based circumference measurements would be a critical improvement.
- **Data Annotation Tools for AI:** The system could serve as an annotation tool for AI development, where clinicians label 3D ultrasound volumes in a natural spatial environment to train segmentation models.

In conclusion, the combination of volumetric ultrasound and the Sony Spatial Reality Display offers a promising shift in how cardiac anatomy is perceived and measured. Despite the hardware challenges encountered, the project provides a solid technical foundation for the further development of interventional visualization tools.

Overview of Generative AI Tools Used

Gemini 3 was used to research the capabilities of Unity´s Universal Renderpipeline. With prompts questioning the options to bypass limitations of using buffer write backs from the fragment shader in the Render Graph, the answers of the LLM guided the decision to implement a scriptable Render Pass to manually bind the buffer to write back measurement marker positions.

Bibliography

[AHIR23] Hiroto Aoki, Yuichi Hiroi, Yuta Itoh, and Jun Rekimoto. Retinal Homing Display: Head-Tracking Auto-stereoscopic Retinal Projection Display. In *29th ACM Symposium on Virtual Reality Software and Technology*, pages 1–10, Christchurch New Zealand, October 2023. ACM.

[KCH22] Dongwoo Kang, Jin-Ho Choi, and Hyoseok Hwang. Autostereoscopic 3D Display System for 3D Medical Images. *Applied Sciences*, 12(9):4288, April 2022.

[KLM+20] Byeol Kim, Yue-Hin Loke, Paige Mass, Matthew R Irwin, Conrad Capeland, Laura Olivieri, and Axel Krieger. A Novel Virtual Reality Medical Image Display System for Group Discussions of Congenital Heart Disease: Development and Usability Testing. *JMIR Cardio*, 4(1):e20633, December 2020.

[Max95] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, June 1995.

[MBK+22] Dharani Maddali, Henrik Brun, Gabriel Kiss, Jon Mikkelsen Hjelmervik, and Ole Jakob Elle. Spatial Orientation in Cardiac Ultrasound Images Using Mixed Reality: Design and Evaluation. *Frontiers in Virtual Reality*, 3:881338, July 2022.

[Son25] Sony Corporation. About Spatial Reality Display - Technical Documentation, 2025. https://xyn.sony.net/en/developer/technical/spatial-reality-display/aboutsrdisplay.

[Uni23] Unity Technologies. Introduction to Scriptable Render Passes, 2023. https://docs.unity3d.com/Packages/com.unity.render-pipelines.universal@14.0/manual/renderer-features/intro-to-scriptable-render-passes.html.

[Uni25] Unity Technologies. Single Pass Stereo Rendering, 2025. https://docs.unity3d.com/Manual/SinglePassStereoRendering.html.

[WDP+19] Gavin Wheeler, Shujie Deng, Kuberan Pushparajah, Julia A. Schnabel, John M. Simpson, and Alberto Gomez. Virtual linear measurement system for accurate quantification of medical images. *Healthcare Technology Letters*, 6(6):220–225, December 2019.