2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 978-1-6654-1298-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISMAR-Adjunct54149.2021.00056

3D Volume Visualization and Screen-based Interaction with Dynamic Ray Casting on Autostereoscopic Display

Ruiyang Li* Tianqi Huang† Hanying Liang‡ Boxuan Han§ Xinran Zhang¶ Hongen Liao‖

Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China

ABSTRACT

Augmented reality (AR) is an emerging technology to improve visualization experiences. However, visualizing volume data is limited in existing AR systems due to the lack of an intuitive and precise exploration scheme. In this paper, we present a 3D augmented volume visualization and screen-based interaction method. An autostereoscopic handheld display is utilized to achieve naked-eye 3D perception, and a stereo camera is adopted to track the display's 6 DoF pose. We implement real-time ray casting with GPU acceleration and enhance the visual experience by defining the dynamic view frustum, clipping interaction and transfer function based on the pose of the handheld display. Our display system achieves real-time rendering and tracking performance with dynamic visual effects, allowing a global overview and detailed visualization of arbitrary clipping planes. We also perform a user study to compare an anatomical landmark annotation task on a 2D environment with our system. The results show a significant reduction in completion time and an improvement in depth perception and comprehension of complex topologies. Furthermore, to illustrate the applicability of our system, we present three volumes from different biological scales.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms Mixed / augmented reality

1 INTRODUCTION

In biological and medical areas, AR has become a practical tool to visualize digital data in physical space [18]. With visual and spatial feedback, users can perceive virtual data as if it appears in physical space. Current applications utilizing AR include presentations of standard mesh models to help novices master anatomical structures [20], the visualization cues for localizing inner targets [27], and the platform for specific tasks such as navigating and planning [13].

Current AR systems utilized in the biomedical domain can be divided into 3 categories, including optical see-through (OST) AR, video-based AR and projection-based AR [10]. Recently, head-mounted displays have drawn much attention [24]. However, wearing the glasses for a long time can cause severe fatigue because of vergence-accommodation conflict [19]. Video-based AR and projection-based AR, on the other hand, mostly present virtual data on 2D display devices or tangible planes [26]. Although these two methods can visualize 3D data, users may encounter visual illusions without correct depth perception, especially for data with complicated structures. In addition to the display system, a majority of existing AR applications can only visualize surface data [12].

*e-mail:li-ry19@mails.tsinghua.edu.cn
†e-mail:huangtianqi@mail.tsinghua.edu.cn
‡e-mail:lianghy20@mails.tsinghua.edu.cn
§e-mail:hbx18@mails.tsinghua.edu.cn
¶e-mail:zhangxinran@tsinghua.edu.cn
‖e-mail:liao@tsinghua.edu.cn

However, there is a great demand for volume visualization, considering the widespread use of volume data in biomedical areas. In brief, users can hardly obtain an intuitive and precise perception of volume data with existing display systems. Glasses-free 3D displays can achieve naked-eye depth perception and allow multiple users. Many technologies have been developed to present multi-view effects. However, these approaches have respective limitations. Light field displays [31] contain time-consuming decomposition calculations, although simulating light fields with high fidelity. Multi-projector displays [33] require one projector per view, making the display systems cumbersome. Holographic displays [34] also have drawbacks, including low frame rate, limited resolution and complicated hardware settings. To guarantee interactive rendering rate and sufficient image resolution, lenticular displays are the preferable choice for portable 3D visualization and interaction system.

To realize specific tasks, AR applications contain the design of interaction with digital data. Hand gesture interaction [3] is widely used in such applications. By designing various hand gestures, the systems can allow various interaction modes [17]. Nevertheless, due to limited operating space, the map relation between large virtual data and hand movement contains large scales, and therefore leads to incorrect distance perception between important landmarks. Screen-based interaction is another mainstream interaction method [26]. By applying interaction based on tangible displays, users can perceive both visual and spatial feedback for intuitive exploration of virtual data.

In this paper, we present an autostereoscopic 3D tablet-based display system with screen-based interaction. The system realizes volume visualization with naked-eye depth perception and intuitive interaction with dynamic rendering effects. We treat the display device as a tangible medium to clip virtual data in physical space. The main contributions of this paper are as follows:

1. We propose a novel way to visualize and interact with 3D volume data by the incorporation of tracked autostereoscopic display and interactive volume rendering.
2. We develop a real-time volume ray casting and screen-based interaction algorithm to achieve dynamic intersecting effects.
3. We perform a user study and prove the enhancement of user comprehension of complex topology by our system.

2 RELATED WORK

2.1 AR with Tablet-based Display Systems

Studies on tablet-based AR systems and applications have emerged. Some applications utilized peripheral display devices at a fixed location to show augmented structures on captured images. Hayashi [14] presented virtual laparoscopic views generated from preoperative CT images and laparoscopic images on two monitors. Bong et al. [4] overlayed anatomical models from preoperative imaging data onto endoscopic images. The augmented view was displayed on a monitor by the side of the operating space. The fixation of monitors causes frequent switches between augmented views and target operation, which increases users' burden.

978-1-6654-1298-8/21/$31.00 ©2021 IEEE
DOI 10.1109/ISMAR-Adjunct54149.2021.00056

By integrating the movement of a tablet-based display device, AR can provide a more intuitive visualization experience. Deng et al. [9] adopted a portable tablet PC to capture surgery scenes and display augmented anatomical structures. Doctors held the display device, and the pose of the display device was obtained by marker-based tracking. Seifert et al. [26] freed users from holding the device by a self-actuated and autonomous display. However, the system configuration limited the portability of tablet-based displays.

The contents of the above display system are limited to surface data. Even though some systems contain the visualization of volumetric data, the 2D display device can hardly provide effective depth perception. Flexpad [28] proposed a projected handheld display to achieve display applications with high degrees of freedom. The research demonstrated an application of curved cross-cuts in volumetric images. However, the displayed content is still an intersected surface of a volume. Chu et al. [8] presented a hybrid rendering of surface and volume data and applied the algorithm to an endoscopic navigation system. Nevertheless, the gradient-based volume shading algorithm highlighted the structure boundary, leading to probable overlooks of internal structures.

2.2 Interaction Method of Autostereoscopic Displays

Interaction methods of autostereoscopic displays include basic interaction by mouse and keyboard, hand gesture and head movement. Peterka et al. [23] utilized head movement as an interaction method. Users could look around a large-scale volume data by changing their head position and orientation. However, frequently changing the head position to view a different part of the data may cause fatigue. Hand gesture interaction is the most common interactive approach in autostereoscopic displays. Alpaslan [1] built a system with two cameras capturing a light source cursor controlled by the user's hand. The movement of the light source was mapped to the transformation of virtual data. Zhang et al. [35] adopted a Kinect camera to track hand location and defined the rotation of the displayed scene according to the hand position. Chen et al. [6] utilized a Leap Motion to recognize hand gestures. They further designed a user study to interact with a rigid heart and a soft liver by a virtual probe.

Although hand gesture interaction is widely used in autostereoscopic displays, it separates the interaction space and data space. Hence, we propose the screen-based interaction method to achieve more intuitive and precise operation.

3 Method

3.1 System Overview

We propose an autostereoscopic 3D tablet-based display system with screen-based interaction to enable users to visualize volume data in physical space, as shown in Fig. 1.

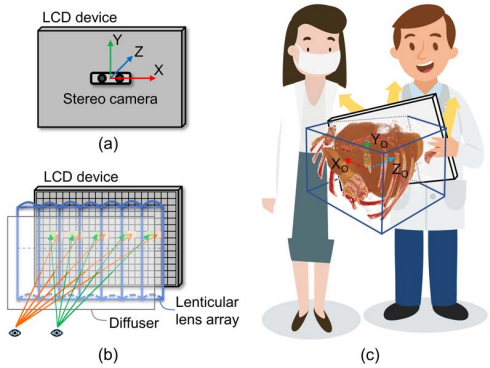

Figure 1: Systematic overview. The back (a) and front (b) sides of the display system. (c) The real-time visualization and interaction stage.

The system hardware includes a stereo camera and an autostereoscopic handheld display device. The stereo camera is attached to the backside of the display device to track the display device's 6 DoF pose. The user flow is composed of an initialization stage and a real-time visualization and interaction stage. During initialization, the user holds the display to determine the fixed location of the virtual volume data in physical space. In the following visualization and interaction stage, the user can clip the volume data with the handheld display and observe different views and appearances by changing the display's 6 DoF in real time. The proposed screen-based interaction and visualization scheme, together with the naked-eye autostereoscopic display, can achieve an effect of exploring the internal information of a real entity and help perceive the topologies of 3D structures and distances between vital landmarks. Moreover, the display provides motion parallax and supports multiple viewers, as well as improving depth perception and user experience.

3.2 Volume Rendering on Autostereoscopic Display

3.2.1 Optical Layout

The utilized autostereoscopic display includes a lenticular lens array, a diffuser and a liquid crystal display (LCD). The lenticular lens array is parallel to the LCD plane. The pixels corresponding to a lenticular lens form an elemental image (EI). As shown in Fig. 2(a), pixels with the same color on the LCD plane are captured by the corresponding camera arranged on a viewing plane. The interval between the adjacent cameras is calculated by $dV = dp \times dis/gap$, where dp is the width of a pixel, gap is the interval between the lenticular lens array and the LCD, and dis is the viewing distance from the lens plane.

The optical parameters can be defined by the given specifications, as shown in Fig. 2(a). The number of views can be calculated by $N_v = pitch/dp$, where $pitch$ is the lenticular lens pitch. The theoretical viewing angle is given by $\alpha = 2\arctan(0.5 \times pitch/gap)$. With a set viewing distance dis, the field of view without image flipping is calculated by $FoV = pitch/gap \times dis$.

3.2.2 Volume Ray Casting Algorithm

The rendering path for surface model displayed on microlens- or lenticular lens-based autostereoscopic displays consists of capturing a multi-viewpoint image array and rearranging the array into elemental images. However, pixels of the identical view are cast along rays separately to get the rendered colors for volume rendering. Therefore, we develop a volume ray casting algorithm for autostereoscopic display. For each pixel on the LCD screen as shown in Fig. 2(b), (c), the related view index i ($i_0 = 0$ for the central view) and the 2D texture coordinate (u,v) (the central coordinate (u_0,v_0) is $(0,0)$) on the ith imaging plane can be calculated [7]. Then, the ray origin $rayO$ and direction $rayD$ can be calculated:

$$rayO(i,u,v) = T_{virtual} Cam_C + tV_i \cdot X' \quad (1)$$

$$rayD(i,u,v) = R_{Y'}(\theta_i) T_{virtual} [(u,v,I_Z)^T - Cam_C] \quad (2)$$

where Cam_C and I_Z are the pre-defined initial 3D location of the central camera and Z-coordinate of the central view image plane. $T_{virtual}$ is the interactive transformation given in Sect. 3.4. X', Y' and O' are original X-axis, Y-axis and Origin (shown in Fig. 2(b)) transformed by $T_{virtual}$. The rotational degree is given by $\theta_i = \arccos[rayD(i_0,u_0,v_0) \cdot (O' - rayO(i,u_0,v_0))]$ and the translational offset is $tV_i = i \times dV$. The rays are cast along their direction and intersect with the bounding box of volume data at points $P_{near} = rayO + t_{near} \cdot rayD$ and $P_{far} = rayO + t_{far} \cdot rayD$ (shown in Fig. 2(c)), where t_{near} and t_{far} are scale lengths of rays. With the ray definition, we cast the rays into the virtual space and sample the volume from P_{near} to P_{far}, as shown in Fig. 2(c). During the sampling process, the accumulated RGBA color after N sampling steps is computed by alpha blending.

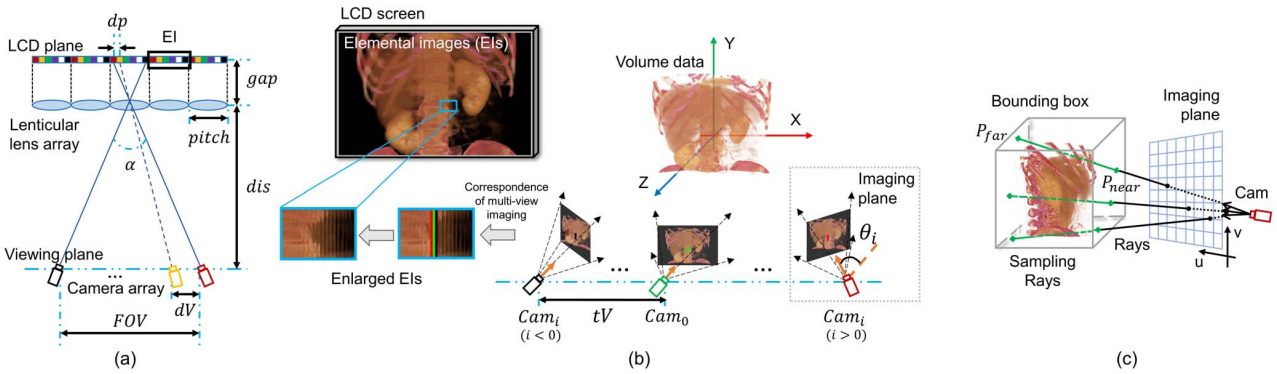

Figure 2: (a) The cross-section view of the autostereoscopic display. (b) The arrangement of multi-view camera array for volume ray casting and the correspondence between multi-view imaging and the pixels on LCD. (c) The enlarged ray casting principle of a single view.

Our proposed algorithm fixes volume data in physical space and casts rays with the virtual camera array moving as the autostereoscopic display, which allows multiple users to explore 3D volume by moving a handheld tangible media and perceive true depth perception, motion parallax and binocular parallax.

3.3 Screen-based Interaction and Dynamic Ray Casting

In this subsection, we exploit the screen pose to define screen-based interaction with dynamic appearances of the rendered volume. The dynamic presence is achieved by dynamic view frustum to mimic the optical illusion [5] in the real world, screen-based clipping, and a dynamic transfer function (TF) scheme.

3.3.1 Dynamic View Frustum for Immersive Display

To mimic the effect that objects appear to be smaller with furtherer distance and also maintain the scale perception while intersecting with the volume, we convert view frustum from perspective (Fig. 3(a)) to orthographic projection (Fig. 3(c)). The distance between the imaging plane and the center of volume data is utilized to control the conversion. This is because the distance between anatomical structures or the size of lesions is vital in surgery, preserving distance length while intersecting with the volume helps surgeons better prepare for clinical operation. Moreover, the design of dynamic view frustum enables viewing the complete picture of the volume at a further screen location.

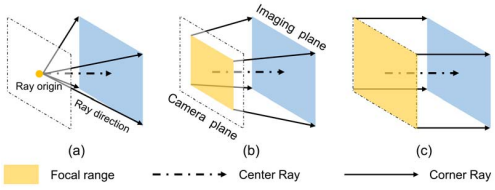

Figure 3: Dynamic view frustum.

3.3.2 Screen-based Clipping Interaction and Dynamic TF

The tangible handheld display is utilized as a clipping medium to interact with the fixed virtual volume data according to the display's pose in physical space, as shown in Fig. 4. After defining the ray origin and direction, the sampling range $\{t_{start} = \max(t_{cr}, t_{near}), t_{end} = t_{far}\}$ is defined to realize screen-based clipping interaction, where t_{near} and t_{far} are the near and far scale length of rays of the intersecting point with the bounding box. Scale length t_{cr} corresponds to the ray's intersecting point P_{cr} with the imaging plane. The design of the clipping interaction has another advantage of considering the characteristics of autostereoscopic displays. Since the performance of depth data is optimal on the focal plane of the lenses [22], clipping the volume at the imaging plane helps provide a better visualization effect for the data plane clipped by the users.

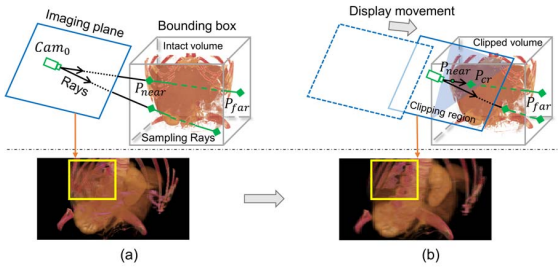

Figure 4: Distance-preserving clipping interaction. (a) No clipping at a far distance. (b) The clipping interaction.

Since the rendered color by ray casting blends an amount of sampled values, a constant TF is insufficient to reveal the complicated 3D structures. After specifying the ray and sampling range, we intend to highlight the vital structures intersected by the imaging plane to better cooperate with the movement of the handheld display.

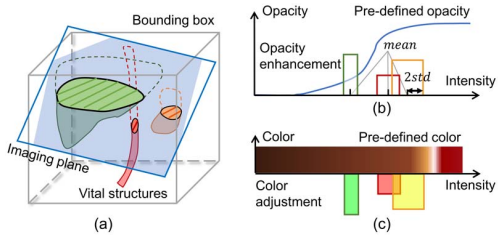

Figure 5: Dynamic transfer function. (a) The imaging plane intersects with pre-segmented vital structures. The opacity values (b) and colors (c) in the $\pm 2std$ range around the $mean$ value of the structures are modified according to the intersected area.

We first input the pre-segmented labels of M kinds of vital structures. For each pixel on the display screen, we track its current position and sample the labels on it. Then, the area of the intersected region (shaded area in Fig. 5(a)) is calculated by summing up all the

samples of the same structure. A normalized intersecting factor of structure s $(0 \leq s < M)$ is computed as follows:

$$intersect_s = \frac{\sum_{(u,v)} sample_s(u,v)}{SA_s \times NF_s} \tag{3}$$

where $sample_s$, SA_s and NF_s represent sampled value, surface area, and normalization factor of structure s. (u,v) is the coordinate on the imaging plane. Based on the pre-defined TF (C_0 and α_0), we define the dynamic RGB color C and opacity α components at greyscale g with the following formulas:

$$\alpha(g) = \alpha_0(g) \times \prod_s [\omega_s(g) + 1] \tag{4}$$

$$C(g) = C_0(g) \cdot \prod_s [\omega_s(g) \cdot C_s + I_{1\times 3}] \tag{5}$$

$$\omega_s(g) = \begin{cases} intersect_s, & if |g - mean_s| \leq 2std_s \\ 0, & else \end{cases} \tag{6}$$

where ω_s is the rectangular window of the structure s with the height defined by $intersect_s$ and the range of $2std_s$ around $mean_s$, as shown in Fig. 5(b). C_s is pre-chose color from HSV color space of structure s as in Fig. 5(c). The overlap of the rectangular windows is because of the similarity between some structures.

Hence, for the voxels belonging to the intersected structures, the opacities are enhanced, and the colors are adjusted to a specific hue. Meanwhile, the larger the intersected area is, the greater the amplitude of the modification will be. Therefore, we modify the appearance of intersected structures by modifying the sampling process.

3.4 Positional Tracking

Positional tracking is a necessary procedure for AR displays. To estimate the 6 DoF pose of the display screen, we utilize a ZED stereo camera attaching to the back of the screen to achieve markerless real-time positional tracking [29] ($T_{virtual}$). The image captured by the camera is also rendered as a background of the displayed volume in order to enhance the immersive experience with visual feedback from the fusion of virtual volume and real scene.

3.5 GPU Accelerated Workflow

The entire program runs three parallel modules: 1) pose tracking module ensures the correct positioning of the multi-view camera array; 2) image rendering module performs ray casting and blending for the displayed image; 3) interaction module achieves interactive volume rendering controlled by screen pose. To obtain real-time performance, the rendering and interaction modules use CUDA architecture to accelerate operating speed. The detailed cooperative workflow of CPU and GPU is shown in supplementary Fig. 1.

4 Experiments and Results

4.1 Implementation

Our software runs on a desktop computer with Intel i9-9900K CPU at 3.6 GHz and 32 GB RAM and NVIDIA GeForce RTX 2080 SUPER. The hardware setup includes a lenticular lens array (with lens pitch 0.991 mm and thickness 2 mm), a diffuser to suppress color Moir effect (with thickness 2 mm), an LCD (with size 12.5 inches, pixel resolution 2560 × 1440 and 235 PPI) and a ZED Mini stereo camera (with resolution 1280 × 720 at 60 frames per second). The display device and attached tracking camera are shown in Fig. 6. We use VTK [25] to import images, CUDA and OpenGL to render elemental images shown on the LCD and ZED SDK to track the screen pose.

The optical parameters of the prototype are listed as follows: the measured viewing angle is 20.5°, the gap is 2.5 mm, the number of views is 9, and therefore the resolution for a single view is 284 × 1440. The field of view is 18.1 cm horizontally (with a viewing distance of 50.0 cm for the handheld device).

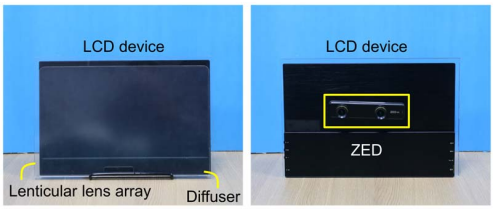

Figure 6: The front (a) and the back (b) sides of the display hardware.

In this and the following section, we present a total of four volumes. The data sources and primary information are listed in Table 1. Abdomen is a CT volume of the abdomen segment from a female with hepatic tumors. Willis is the Willis' circle in the human brain imaged with time-of-flight (ToF) magnetic resonance angiography (MRA). Neurons is a post-integrated light microscopy capture of three neurons. Body is a low-resolution CT capture of the whole body from a healthy male.

Table 1: Volume data sources and basic information

Name	Modality	Dimension	Source
Abdomen	CT	$512 \times 512 \times 129$	[16]
Willis	TOF-MRA	$560 \times 560 \times 140$	[30]
Neurons	Light Microscopy	$1024 \times 1024 \times 119$	[15]
Body	CT	$512 \times 512 \times 521$	[21]

4.2 Display Performance

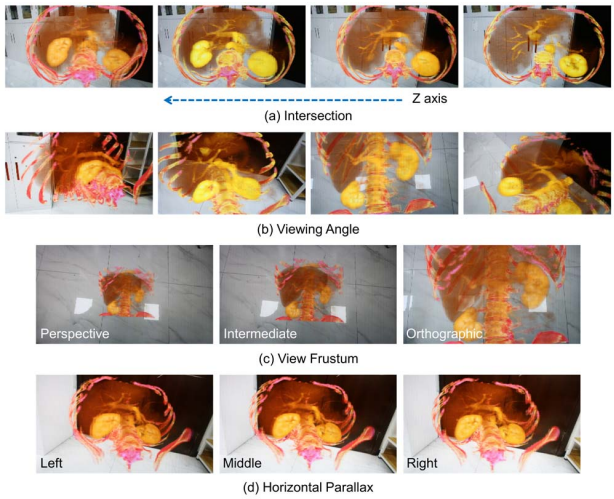

Figure 7: Display performance. (a) The effects of clipping interaction. (b) Dynamic rendering effects from different views. (c) Dynamic view frustums. (d) Horizontal parallax of the autostereoscopic display.

We changed the distance between the display device and the origin of the volume Abdomen and adjusted the display pose to achieve different clipping planes and rendering effects. The display performance of dynamic ray casting is shown in Fig. 7. Fig. 7(a) presents the clipping interaction along the axial direction. When the clipping plane intersects with key organs such as vessels, the opacities of

corresponding greyscales are enhanced and color components are adjusted. Fig. 7(b) further provides rendering effects from different viewing angles. When we moved the display device from far to near, the volume appeared as a scaling effect. As shown in Fig. 7(c), the dynamic view frustum allows a global overview and a detailed visualization by changing the display device position. Besides, our display system realizes naked-eye autostereoscopic display with a lenticular lens array, providing horizontal parallax as shown in Fig. 7(d).

4.3 Evaluation

4.3.1 Rendering Speed

With regard to quantitative performance, we calculated the mean rendering speed in one minute for different volumes, during which we interacted with the data by changing the display pose. The results are 28.43 fps, 27.33 fps, 26.77 fps and 32.23 fps for Abdomen, Willis, Neurons and Body, respectively. Therefore, our system realizes real-time rendering performance at around 30 fps for various datasets, satisfying the requirements of real-time AR applications.

4.3.2 User Study

To validate the effectiveness of the proposed display system, we designed a control experiment to compare the user performance of an annotation task on our system ('3D' for short) versus the conventional tools of the 2D display device and 3D Slicer software [11] ('2D' for short). The users were all novices in clinical studies and unfamiliar with the used anatomical structures. We chose the Willis' circle as the target anatomical structure since the vessel topology is complicated for novices. On the 3D display device, users operated in 3D space and visualized the volume data from different viewpoints by changing the screen pose, as shown in Fig. 8(a). When the anatomical landmark was on the clipping plane, the user marked the target by pointing it out. On the 2D display device, users could utilize the slicing tool, ROI selection, and volume rendering window in 3D Slicer software to accomplish the annotation task with mouse and keyboard interaction, as shown in Fig. 8(b). The completion time to mark two anatomical landmarks (the anterior communicating artery and the top of the basilar artery) was recorded. Users were also asked to fill in a questionnaire throughout the experiment. The evaluation was based on the objective completion time and subjective feedback.

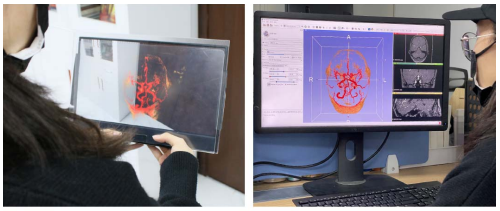

Figure 8: The annotation task for clinical novices to accomplish with our system (a) and the conventional 2D environment (b).

The procedures of the user study were as follows: (1) Users were first asked about their familiarity with Willis' circle and 3D Slicer and their former experience with AR and VR devices. (2) Users then briefly learned about the anatomical structure of Willis' circle and confirmed the two landmarks to annotate. (3) Users could practice on the annotating environment for a few minutes before the formal annotation task and completed the task on two display devices, respectively. After each trial, users scored the subjective ratings on the corresponding display and interaction method.

14 participants took part in the evaluation (9 males and 5 females, with a mean age of 25). All participants except one were unfamiliar with the structure of Willis' circle. 4 participants were proficient in using 3D Slicer, and 9 were familiar with the basic usage. 12 participants reported experiences with AR or VR devices. A specialist was responsible for justifying whether the participants correctly found the landmarks and the completion time was recorded. After each trial on different display devices, the participants provided subjective ratings on the ease of labeling interaction (How convenient is it to label a landmark?), depth perception (How well do you sense the depth information?) and the improved understanding of the vessel topology (How well does the visualization improve your understanding of the topology?). The questions were designed as a ten-point Likert scale.

The results of the objective completion time and the subjective ratings were shown in Fig. 9. The completion time was 63.14 ± 25.62 s for the 2D display device and 23.57 ± 11.27 s for our 3D display device as shown in Fig. 9(a). We used a one-way ANOVA (Analysis of Variance) to examine the significant reduction of completion time with our system. The analysis results showed significant differences in completion time between the two methods with $F(1,13) = 25.10$, $p < 0.001$ at the significance level of 0.05. The average scores and the distribution of the subjective ratings were shown in Fig. 9(b) and Fig. 9(c). The subjective ratings of ten-point Likert scales were analyzed by the Wilcoxon signed-rank test [32]. Significant differences appeared in perception (W = 80.5, $p = 0.0105$) and understanding (W = 62, $p = 0.0078$), while labeling (W = 70.5, $p = 0.0889$) failed to reject the null hypothesis.

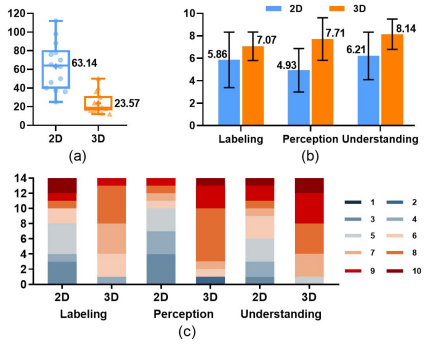

Figure 9: Evaluation results of the user study. (a) Completion time. The box and whiskers show the minimum, maximum, first and third quartiles, with median (bands) and mean ('+' and numbers) values marked. (b) Average scores and standard deviations of subjective ratings. (c) The distribution of subjective ratings.

The above statistical analysis demonstrates that our display system and rendering algorithm provide a better perception, ease the difficulty of understanding and shorten the time to perform specific tasks on complicated topologies on the whole, which is especially helpful for novices in clinical areas.

5 Applications

In this section, we present three visualization scenarios: Neurons, Willis and Body, as shown in Fig. 11. Screen-based clipping interaction can help researchers understand the neuromorphic information and junction (Fig. 10(a)) with better perception. With our system, novices can also interact with the vessel structures in physical space and clip along vessels to observe blood flow direction in Fig. 10(b). We can also visualize the imaging data of a whole body, as shown in Fig. 10(c). Doctors can slide through the data in physical space to help perform systemic diagnosis or preoperative planning. In conclusion, our system can display volume data in different biological scales, from a single cell through organs to a living body.

Our system helps relieve users' burden of figuring out complicated structures and improve the outcomes of certain tasks.

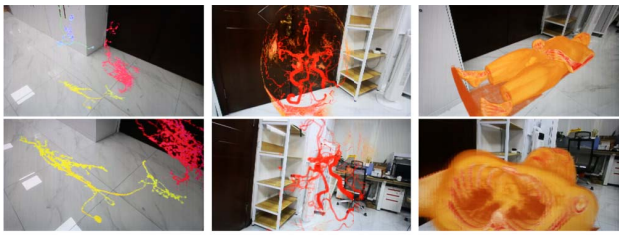

Figure 10: Multi-view visualization of volume data from different biological scales: Neurons (a), Willis (b) and Body (c).

6 Conclusions

To summarize, we realized real-time volume visualization and screen-based interaction on an autostereoscopic handheld display device with a rendering speed of around 30 fps. The screen-based interaction method provided dynamic rendering effects to highlight the intersecting anatomical structures. The screen-based clipping interaction in an arbitrary direction helped increase the degree of freedom to view internal structures of the volume data. We also performed a user study to validate the effectiveness of our system. Results showed significant improvements in depth perception and understanding of complicated structures.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (82027807, 81771940), Beijing Municipal Natural Science Foundation (7212202), and Tsinghua University Initiative Scientific Research Program (20197020014).

References

[1] Z. Alpaslan and A. Sawchuk. *Three-dimensional interaction with autostereoscopic displays*, volume 5291 of *Electronic Imaging 2004*. SPIE, 2004.

[2] M. Billinghurst. Hands in space gesture interaction with augmented-reality interfaces.

[3] J. H. Bong, H. J. Song, Y. Oh, N. Park, H. Kim, and S. Park. Endoscopic navigation system with extended field of view using augmented reality technology. *Int J Med Robot*, 14(2), 2018.

[4] M. A. Changizi, A. Hsieh, R. Nijhawan, R. Kanai, and S. Shimojo. Perceiving the present and a systematization of illusions. *Cognitive Science*, 32(3):459–503, 2008.

[5] G. Chen, T. Huang, Z. Fan, X. Zhang, and H. Liao. A naked eye 3d display and interaction system for medical education and training. *Journal of biomedical informatics*, 100:103319, 2019.

[6] G. Chen, C. Ma, Z. Fan, X. Cui, and H. Liao. Real-time lens based rendering algorithm for super-multiview integral photography without image resampling. *IEEE transactions on visualization and computer graphics*, 24(9):2600–2609, 2017.

[7] Y. Chu, X. Li, X. Yang, D. Ai, Y. Huang, H. Song, Y. Jiang, Y. Wang, X. Chen, and J. Yang. Perception enhancement using importance-driven hybrid rendering for augmented reality based endoscopic surgical navigation. *Biomed Opt Express*, 9(11):5205–5226, 2018.

[8] W. Deng, F. Li, M. Wang, and Z. Song. Easy-to-use augmented reality neuronavigation using a wireless tablet pc. *Stereotactic and Functional Neurosurgery*, 92(1):17–24, 2014.

[9] M. Eckert, J. S. Volmerg, and C. M. Friedrich. Augmented reality in medicine: Systematic and bibliographic review. *JMIR Mhealth Uhealth*, 7(4):e10967, 2019.

[10] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, and M. Sonka. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[11] K. A. Gavaghan, M. Peterhans, T. Oliveira-Santos, and S. Weber. A portable image overlay projection device for computer-aided open liver surgery. *IEEE Trans Biomed Eng*, 58(6):1855–64, 2011.

[12] J. Hallet, L. Soler, M. Diana, D. Mutter, T. F. Baumert, F. Habersetzer, J. Marescaux, and P. Pessaux. Trans-thoracic minimally invasive liver resection guided by augmented reality. *J Am Coll Surg*, 220(5):e55–60, 2015.

[13] Y. Hayashi, K. Misawa, M. Oda, D. J. Hawkes, and K. Mori. Clinical application of a surgical navigation system based on virtual laparoscopy in laparoscopic gastrectomy for gastric cancer. *Int J Comput Assist Radiol Surg*, 11(5):827–36, 2016.

[14] A. Institute. Neuron image data sets.

[15] IRCAD. 3d-ircadb-01 database.

[16] A. Karambakhsh, A. Kamel, B. Sheng, P. Li, P. Yang, and D. D. Feng. Deep gesture interaction for augmented anatomy learning. *International Journal of Information Management*, 45:328–336, 2019.

[17] M. Kersten-Oertel, P. Jannin, and D. L. Collins. The state of the art of visualization in mixed reality image guided surgery. *Comput Med Imaging Graph*, 37(2):98–112, 2013.

[18] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of imaging science and technology*, 53(3):30201–1–30201–14, 2009.

[19] S. K. Motsinger. Complete anatomy. *Journal of the Medical Library Association*, 108(1), 2020.

[20] N. L. of Medicine. Visible human project.

[21] J.-H. Park, S.-W. Min, S. Jung, and B. Lee. Analysis of viewing parameters for two display methods based on integral photography. *Applied Optics*, 40(29):5217–5232, 2001.

[22] T. Peterka, R. Ross, H. Yu, K.-L. Ma, R. Kooima, and J. Girado. *Autostereoscopic display of large-scale scientific visualization*, volume 7237 of *IS and T/SPIE Electronic Imaging*. SPIE, 2009.

[23] R. Rahman, M. E. Wood, L. Qian, C. L. Price, A. A. Johnson, and G. M. Osgood. Head-mounted display use in surgery: A systematic review. *Surg Innov*, 27(1):88–100, 2020.

[24] W. J. Schroeder, L. S. Avila, and W. Hoffman. Visualizing with vtk: a tutorial. *IEEE Computer graphics and applications*, 20(5):20–27, 2000.

[25] J. Seifert, S. Boring, C. Winkler, F. Schaub, F. Schwab, S. Herrdum, F. Maier, D. Mayer, and E. Rukzio. Hover pad: interacting with autonomous and self-actuated displays in space. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 139–147, 2014.

[26] M. Solbiati, K. M. Passera, A. Rotilio, F. Oliva, I. Marre, S. N. Goldberg, T. Ierace, and L. Solbiati. Augmented reality for interventional oncology: proof-of-concept study of a novel high-end guidance system platform. *Eur Radiol Exp*, 2:18, 2018.

[27] J. Steimle, A. Jordt, and P. Maes. *Flexpad: highly flexible bending interactions for projected handheld displays*, page 237–246. Association for Computing Machinery, 2013.

[28] Stereolabs. Zed mini stereo camera - stereolabs.

[29] K. Timmins. Aneurysm detection and segmentation challenge.

[30] G. Wetzstein, D. Lanman, M. Hirsch, W. Heidrich, and R. Raskar. Compressive light field displays. *IEEE computer graphics and applications*, 32(5):6–11, 2012.

[31] F. Wilcoxon. *Individual comparisons by ranking methods*, pages 196–202. Springer, 1992.

[32] X. Xia, X. Zhang, L. Zhang, P. Surman, and Y. Zheng. Time-multiplexed multi-view three-dimensional display with projector array and steering screen. *Opt Express*, 26(12):15528–15538, 2018.

[33] F. Yaras, H. Kang, and L. Onural. State of the art in holographic displays: A survey. *Journal of Display Technology*, 6(10):443–454, 2010.

[34] J. Zhang, X.-Q. Xu, J. Liu, L. Li, and Q.-H. Wang. Three-dimensional interaction and autostereoscopic display system using gesture recognition. *Journal of the Society for Information Display*, 21(5):203–208, 2013.