Integral 3D Display of Real Objects Based on Multi-view Image Acquisition Using 3D Gaussian Splatting

Michitaka Ono[1], **Naoto Matsugi**[1], **Masahiro Kawakita**[1], **Norihiko Kawai**[1]

norihiko.kawai@oit.ac.jp
[1]Faculty of Information Science and Technology, Osaka Institute of Technology
1-79-1 Kitayama, Hirakata, Osaka, 573-0196 Japan
Keywords: Integral 3D, Lens array, 3D Gaussian Splatting, Novel view synthesis, Stereoscopic image.

ABSTRACT

We propose a method for generating integral 3D images using only a single consumer camera and 3D Gaussian Splatting. From casually captured multi-view images, we reconstruct a radiance field and generate virtual-view images. These are converted into elemental images for integral 3D displays, enabling realistic and low-cost 3D visualization of real objects.

1 Introduction

Realistic visualization is crucial in e-commerce and virtual exhibitions. However, conventional product images captured from limited viewpoints fail to represent shape and material. For this problem, integral 3D display technology has been developed that allows users to view 3D representations of objects with naked eyes. Conventionally, 3D contents viewable on such displays have been generated from images captured by multi-camera systems [1] or lens arrays [2]. However, these require complex equipment and calibration to capture images of target objects. We propose a simplified solution: combining 3D Gaussian Splatting (3DGS) [3] with integral 3D display technology using only a handheld camera.

Our main contributions are as follows:

- A novel pipeline integrating 3DGS and integral imaging to synthesize realistic stereoscopic displays from casually captured images using a single consumer-grade camera.
- The ability to visualize real objects of various sizes from small items to large scenes while preserving material properties such as gloss, softness, or transparency.
- Virtual camera flexibility within the reconstructed 3D scene, allowing for adjustable parallax and depth emphasis.

2 Proposed Method

2.1 Virtual Space Generation via 3DGS

Objects are photographed from various viewpoints without requiring special rigs or calibration. Camera poses and 3D points are estimated using COLMAP [4], which performs Structure-from-Motion. These are used to train the 3DGS model, which assigns 3D Gaussians in space, thus constructing a radiance field that mimics reality.

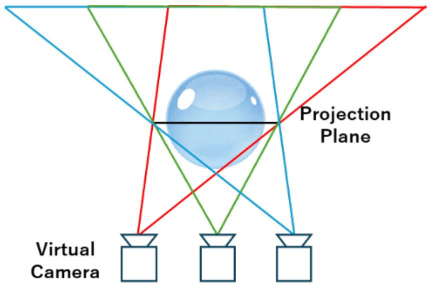

Fig. 1 Lens-shift method

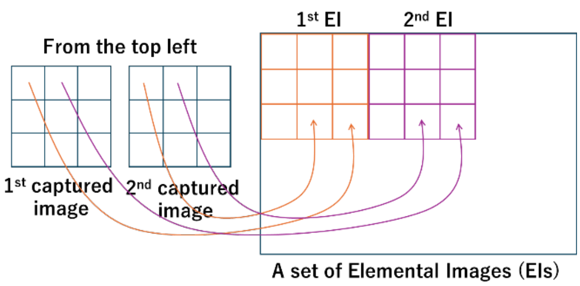

Fig. 2 Generation of elemental images

2.2 Multi-view Image Capture in Virtual Space

We perform systematic viewpoint capture in the virtual scene using a raster-scan-like camera movement. The camera translation interval is predefined based on the desired parallax to be expressed, and the same interval is applied uniformly in both horizontal and vertical directions for consistent sampling. The virtual camera movement follows a lens-shift method, where the imaging plane remains fixed while the viewpoint shifts in parallel, as shown in Fig. 1. The number of viewpoints and image resolution are determined based on the lens array geometry and display resolution.

2.3 Elemental Image Generation

Captured images are rearranged to form a set of elemental images corresponding to a single lens array. Specifically, from the captured images, each pixel is extracted one by one following a raster scan from the top-left to the bottom-right. These pixels are then reassigned to match the microlens layout by pasting them in reverse raster order, from the bottom-right to the top-left direction, as shown in Fig. 2. The image is resized and cropped to match the actual display resolution and lens alignment.

2.4 Stereoscopic Image Display

The final elemental image is shown on a high-resolution display behind a lens array. Each lens projects rays from different pixels in different directions, enabling the viewer to perceive a full-parallax 3D image with depth.

3 Experiments

In this study, we investigated the effectiveness of the proposed method by generating elemental images from multi-view captures of real objects with varying material properties and evaluating them through an integral 3D display. The experimental setup included a display and a lens array, whose specifications are summarized in Tables 1 and 2, respectively.

Three types of real objects with various materials and scales were selected for evaluation. For each object, we captured more than 300 images from multiple viewpoints and used them to train the 3DGS model. To facilitate batch capture, we extended the open-source 3DGS renderer “Splatapult” [5] by implementing camera control and image capture automation.

The initial camera pose in the virtual scene was chosen from among the camera poses estimated by COLMAP. Each rendered view was generated at a resolution of 197×111 pixels. To match the display configuration, we synthesized 20×20 sub-views per microlens.

Since the resulting elemental image exceeded the native 4K display resolution (3840×2160 pixels), we applied a scaling factor of 0.983 to ensure alignment with the microlens pitch, and cropped the image to fit within the display boundaries.

The test objects included: (1) a small object exhibiting strong reflection and refraction, (2) a plush object with a soft, fibrous surface structure, and (3) a relatively large indoor scene in a university building. For each case, we varied the camera sweep range and adjusted the depth position of the projection plane to observe its impact on the perceived 3D appearance. The following subsections show the experimental results in detail.

Table 1 Specifications of display

Display type	OLED
Size	13.3 inches
Resolution	4K (3840×2160 pixels)

Table 2 Specifications of lens array

Size	294 mm × 165 mm
Thickness	5 mm
Lens shape	Spherical
Microlens size	1.5 mm square
Lens arrangement	Square grid
Radius of curvature	1.820 mm
Focal length	3.699 mm
Viewing angle	23 degrees
Material	PMMA
Refractive index	1.492

Fig. 3 Captured views in Experiment 1

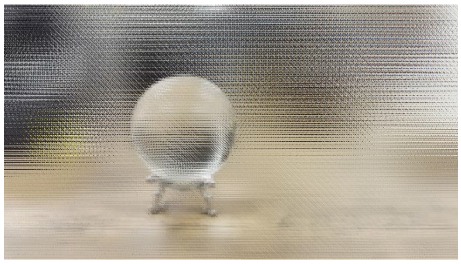

Fig. 4 A set of elemental images with 23° viewing angle in Experiment 1

Fig. 5 Views through lens array from Fig. 4

3.1 Experiment 1 with a Reflective and Refractive Object

As an object exhibiting reflection and refraction, we used a glass ball. A total of 328 images were used for 3DGS training. The camera was moved to cover a viewing angle of approximately 23°, and a subset of the captured views is shown in Fig. 3. The glass ball itself remains stationary while the background shows parallax.

The resulting set of elemental images, shown in Fig. 4, was generated with the projection plane aligned to the contour of the glass ball. As a result, both the background objects and the objects visible through the glass ball appear blurred in the elemental image. However, as shown in Fig. 5, when viewed through the lens array, the image reveals both the background and foreground clearly, and vertical and horizontal parallax is perceivable with slight viewpoint shifts.

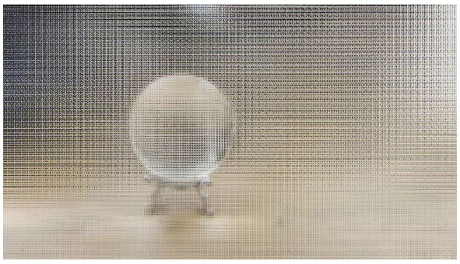

Fig. 6 A set of elemental image with 46° viewing angle in Experiment 1

Fig. 7 Views through lens array from Fig. 6

Next, we increased the sweep range of the virtual camera to double the viewing angle (approximately 46°), keeping the projection plane fixed. The resulting set of elemental images is shown in Fig. 6 and the lens-array view in Fig. 7. The larger disparity led to increase blur in the elemental image, but when viewed through the lens array, we could recognize the objects. While some improvement in depth perception was observed, the enhancement was not as clear as initially expected.

We then repositioned the projection plane to a location between the glass ball and the background objects. In the set of elemental images shown in Figs. 8, both the glass ball and the background objects appeared blurred. When viewed through the lens array, as shown in Fig. 9, the glass ball seemed to come forward and move side to side horizontally in response to viewpoint changes, creating an experience similar to observing the object from different angles.

3.2 Experiment 2 with a Soft, Fibrous Object

For the second experiment, we selected a stuffed toy with a soft and fibrous surface texture. A total of 369 images were captured and used to train the 3DGS model. The projection plane was set to be aligned to the contour of the toy, and the virtual camera was swept across a range corresponding to a viewing angle of approximately 23°. Fig. 10 shows a subset of the captured images.

The set of elemental images generated from the rendered views is shown in Fig. 11, and the reconstructed 3D views are presented in Fig. 12. Because the projection plane was aligned to the toy's contour, the contour itself appears sharp in the elemental image, while the background objects and the parts located in front of the

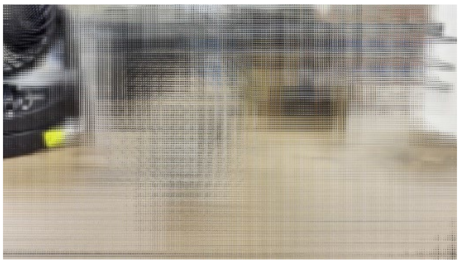

Fig. 8 A set of elemental images with projection plane position changed in Experiment 1

Fig. 9 Views through lens array from Fig. 8

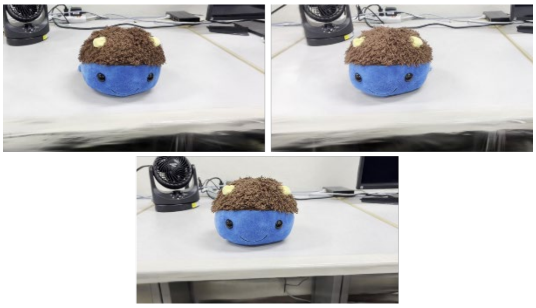

Fig. 10 Captured views in Experiment 2

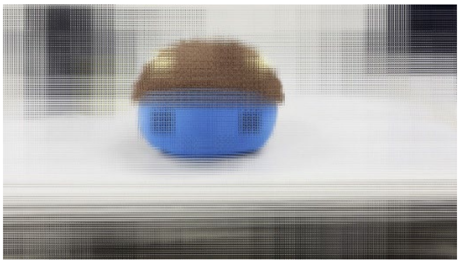

Fig. 11 A set of elemental images in Experiment 2

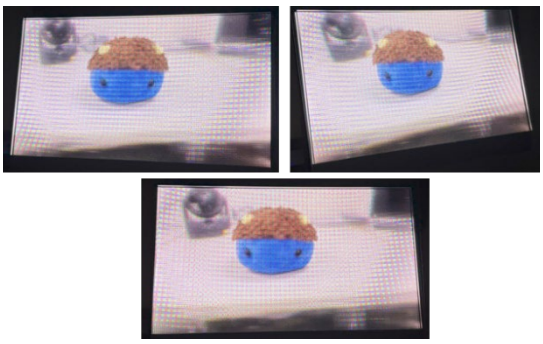

Fig. 12 Views through lens array from Fig. 11

contour - such as the eyes and horns - appear blurred. However, when viewed through the lens array, the 3D shape of the toy is perceived clearly from multiple angles, and the texture of the fibrous surface is also conveyed effectively. These results demonstrate that the proposed method is capable of reproducing not only geometric depth but also subtle material characteristics in soft-textured objects.

3.3 Experiment 3 with a Large Indoor Scene

For the third experiment, we selected a large-scale real-world scene - specifically, the entrance area of the Hirakata Campus at Osaka Institute of Technology. A total of 312 images were captured and used to train the 3DGS model. The virtual camera was swept across a range corresponding to a viewing angle of approximately 23°, and a subset of the captured images is shown in Fig. 13.

The set of elemental images generated from the reconstructed scene is shown in Fig. 14, and the resulting 3D views through the lens array are shown in Fig. 15. In this case, the projection plane was set around the monitor located near the center of the image. As a result, foreground elements such as the handrails and stair edges appear blurred in the elemental image. When viewed through the lens array, the handrails still slightly blur but appear in front of the monitor. The large disparity of the handrails makes them appear to come out toward the viewer, enhancing the overall sense of depth.

3.4 Discussion

In the three cases, 3D visualization was successfully achieved through the lens-array display. The shape and material of the objects were effectively conveyed through the generated elemental images.

However, in some scenes - particularly the indoor scene or those involving wide spatial coverage - the quality of the 3DGS reconstruction was limited and has a negative impact on the perceived quality of the integral 3D display. When the reconstruction is inaccurate, the synthesized views contain artifacts, which in turn degrades the realism.

Furthermore, although our experiments subjectively confirmed the effects of the projection plane and viewing angle on 3D perception, we need to quantitatively evaluate these effects through user studies. Additionally, dynamic adjustment mechanisms according to user preferences should be explored to improve perceptual comfort and interactivity.

4 Conclusion

We proposed a method that combines 3D Gaussian Splatting with integral imaging to enable 3D visualization from simple image capture. This system provides a low-cost, intuitive 3D display method, adaptable to various object types and usable in consumer or exhibition contexts. Future work includes further investigation of the effects of each parameter on 3D perception and implementation of dynamic adjustment functions.

Fig. 13 Captured views in Experiment 3

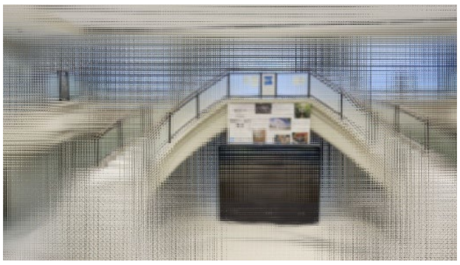

Fig. 14 A set of elemental images in Experiment 3

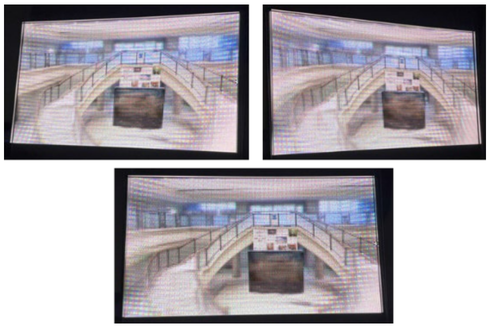

Fig. 15 Views through lens array from Fig. 14

Acknowledgement This work was supported by JSPS KAKENHI Grant Number JP25H01154, JST SCORE (JPMJST2051), and Hoso Bunka Foundation.

References

[1] K. Ikeya, K. Hisatomi, M. Katayama and Y. Iwade, “Integral 3D Contents Production from Multi-View Images: 3D Modeling and 3D Image Conversion of Sports Scene,” The Journal of the Institute of Image Information and Television Engineers, Vol. 67, No. 7, pp. J229-J240, (2013). (In Japanese)

[2] S. Kono, M. Kameyama and M. Kawakita, “Aerial-imaging Light-field Camera and Integral 3D Video Communication System,” Proc. Optica Imaging Congress 2024, DTh5F.5, (2024).

[3] B. Kerbl, G. Kopanas, T. Leimkuehler and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” ACM Transactions on Graphics, Vol. 42, No. 4, pp. 139:1-139:14 (2023).

[4] J.L. Schönberger, J.-M. Frahm, “Structure-from-Motion Revisited,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, (2016).

[5] “Splatapult,” https://github.com/hyperlogic/splatapult, (Accessed on 18/06/2025).