3DSAp2-13L
Late-News Paper

Enhancing Image Quality in Directional Volumetric Displays with Red-Net

Seigou Takeshita[1], Takashi Nishitsuji[2], Naoto Hoshikawa[3], Hirotaka Nakayama[4], Tomoyoshi Shimobaba[1], Tomoyoshi Ito[1], Atsushi Shiraki[1]

seigou_0713@chiba-u.jp
[1] Graduate School of Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan.
[2] Faculty of Science, Toho University, 2-2-1 Miyama, Funabashi-shi, Chiba 274-0851, Japan.
[3] Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki 305-8577, Japan.
[4] Center for Computational Astrophysics, National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan.
Keywords: volumetric display, deep learning, enhancing image quality

ABSTRACT

The existing directional volumetric display, which employs threads and a projector, exhibits a significant problem: noise within the displayed image when juxtaposed with the original image. Herein, we harnessed deep learning techniques to expedite the process of enhancing image quality in contrast to conventional methodologies.

1 Introduction

Because humans inhabit a three-dimensional world, three-dimensional information holds great significance. Among these aspects, interest in three-dimensional images and video is steadily increasing. Consequently, there has been a surge in research dedicated to three-dimensional displays for projecting such images. Various three-dimensional display technologies have emerged. Examples include holography [1], a method that captures data from multiple light sources with different wavelengths to produce a three-dimensional visual effect, and volumetric displays [2], where the display possesses volume and allows viewing from any direction in 360°.

This study employs a directional volume display [3], which is capable of projecting distinct images in various directions by directly rendering them onto a three-dimensional surface. Owing to its artistic and directional attributes, which are different from those of conventional displays, this technology is anticipated to find applications in encryption and media art.

Several types of directional volumetric displays were employed, including 3D crystals, light-emitting diode (LED) configurations, and the integration of threads with a projector, as illustrated in Figs. 1, 2, and 3, respectively. The projection results indicate that varying images are exhibited depending on the observation direction.

Directional volumetric displays exhibit more noise in the displayed image than in the original image.

Previously, an image quality improvement algorithm employing the successive approximation method was introduced in response to this challenge [5].

This algorithm compares the displayed image on the volume display with the original image and subsequently updates the input image to enhance image quality. However, with each update of the input image, the displayed image requires recalculation, resulting in a time-consuming process for image quality enhancement. When computational costs are high, achieving an adequate frame rate for real-time projection becomes challenging. We propose an image quality improvement algorithm that employs deep learning to address this issue. This approach discerns the relationship between the pixel values in the original and displayed images and uses it to update the input image. This study aims to enhance image quality while reducing computational costs to attain a satisfactory frame rate.

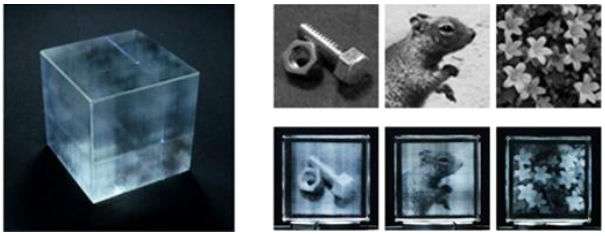


Fig. 1 Directional volumetric display using 3D crystals.

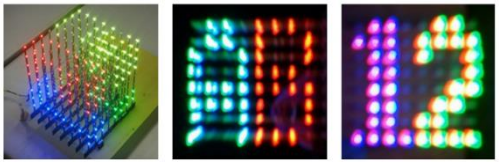


Fig. 2 Directional volumetric display using LED.

Fig. 3 Directional volumetric display using threads and a projector.

2 Previous Research

The directional volumetric display using threads and a projector generates voxel data from multiple distinct images. Voxel data, in essence, represent three-dimensional information and is depicted on the thread-based-volume display, resembling the rectangular object illustrated in Fig. 4.

In figure 4, O_1 and O_2 represent the original images intended for display, with $O_1(u_1, v_1)$ and $O_2(u_2, v_2)$ denoting the pixel values of these images, respectively. Equation (1) allows us to derive the voxel data $V(x, y, z)$ at coordinates $(X, Y, Z) = (x, y, z)$ from these values. However, the display images produced using this algorithm [5] tend to exhibit low contrast and diminished brightness.

$$V(x, y, z) = \lambda\{O_1(u_1, v_1) + O_2(u_2, v_2)\}. \quad (1)$$

In equation (1), λ is a constant normalizing voxel value between 0 and 255. Voxels are summed along each axis direction in a volumetric display to generate output images. The pixel value at each image coordinate can be calculated using equations (2) and (3).

$$D_1(u_1, v_1) = \sum_{w_1} V(x, y, z). \quad (2)$$

$$D_2(u_2, v_2) = \sum_{w_2} V(x, y, z). \quad (3)$$

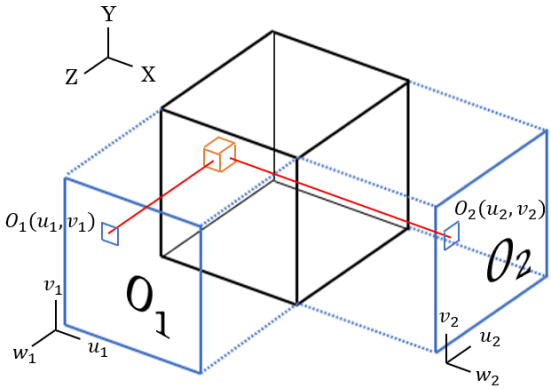


Fig. 4 Overview of directional volumetric display using threads and a projector.

Displayed images on a directional volumetric display often exhibit significant noise compared to the original images. An image quality improvement algorithm based on successive approximation was introduced to address this issue.

This process involves comparing the output image, computed from voxel data, with the original image and subsequently updating the input image to enhance the quality of the final displayed image. Equation (4) is employed to update the input image.

$$I_i(u_i, v_i)^{k+1} = I_i(u_i, v_i)^k + (O_i(u_i, v_i) - D_i(u_i, v_i)^k). \quad (4)$$

In equation (4), k represents the number of iterations, $I_i(u_i, v_i)^k$ denotes the input image for the k th iteration, $D_i(u_i, v_i)^k$ corresponds to the displayed image for the k th iteration, and $O_i(u_i, v_i)$ represents the original image.

The final $I_i(u_i, v_i)^{k+1}$ is utilized to derive voxel data and produce displayed images, ultimately enhancing image quality.

3 Proposed method

This section elucidates the algorithms and network structure employed to enhance image quality through deep learning.

3.1 Overview of image quality improvement

In this section, we explore methods for enhancing image quality when displaying grayscale images on the front, side, and top surfaces. Each source image, O_1,O_2, and O_3, has dimensions of $w \times h$. For simplicity, we consider the case of grayscale images with a channel in each original image. We calculate the voxel data from these three images using equation (5).

$$V(x, y, z) = \lambda \sum_{i=1}^{3} O_i(u_i, v_i). \quad (5)$$

Normalize the obtained voxel data to a value between 0 and 255 using equation (6).

$$V(x, y, z) = \begin{cases} 0 & (V(x, y, z) < 0) \\ V(x, y, z) & (0 < V(x, y, z) < 255) \\ 255 & (0 < V(x, y, z)) \end{cases}. \quad (6)$$

The output images D_1,D_2, and D_3in the front and side directions are obtained from the normalized voxel data using equation (7).

$$D_i(u_i, v_i) = \sum_{W_i}^{3} V(x, y, z). \quad (7)$$

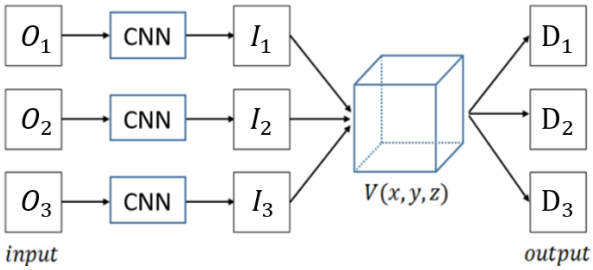
The obtained D_i is employed as input, and O_i serves as the supervisory data for training a convolutional neural network (CNN) model. The model learns the pixel variations between the original and output images to minimize the differences between their pixel values. As defined in equation (8), the mean squared error shown is used as the error function $E(w)$ during the model training process.

$$E(w) = \frac{1}{wh} \sum_{i=1}^{2} \sum_{j=0}^{w-1} \sum_{k=0}^{h-1} \left(D_i(j, k) - O_i(j, k)\right)^2. \quad (8)$$

Figure 5 provides an overview of the proposed method using the CNN learned in the previous procedure.

The original images, denoted as O_i, are fed into the CNN, producing the resulting image I_i , which is subsequently used to derive voxel data and generate displayed images. This process aligns with the conventional approach of updating input images by incorporating differences in pixel values between the original and output images, as seen in conventional

image quality improvement algorithms. The updated input image, I_i is the dataset from which the features suitable for the algorithm in equation (5) are extracted. It also encapsulates changes in pixel values occurring during the output image computation. The voxel data obtained from



I_i are expected to enhance the image output quality via equation (7).

Fig. 5 Overview of image quality Improvement using CNN.

3.2 Data set creation

Three images were randomly selected from CIFAR-10, comprising monochrome images with 32 × 32 pixels. Using the equation from Section 3.1, output images were computed, resulting in a dataset comprising 9,999 pairs of original and output images.

3.3 Network structure

The CNN model employed in this study is based on residual encoder-decoder networks (RED-Net), as described in the literature [5]. This model comprises six convolutional layers in the first half and six deconvolutional layers in the second half, with the ReLU activation function being used along with three introduced skip connections [6]. Refer to Fig. 6 for an illustration.

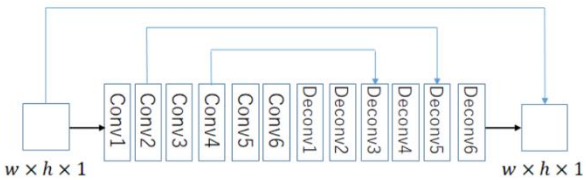


Fig. 6 Overview of network structure.

3.4 Deep Learning model building and execution

A deep learning model is built using the dataset, where the output images obtained from the voxel data serve as input, and the original images as output.

This research will use Tensorflow, a deep learning framework with a batch size of 32, 100 epochs, a learning rate of 0.0001, and the adaptive moment estimation (ADAM) optimization method.

4 Results and discussion

4.1 Image quality improvement results

To demonstrate the effectiveness of the proposed method, we compare images obtained with no image quality improvement, a conventional image quality improvement algorithm with one iteration and the proposed method. We evaluated the images using structural similarity (SSIM). The results from the proposed method are generated using equations (5), (6), and (7) with the data obtained by inputting images into the trained network. Table 1 displays the SSIM results for each method, and Fig. 7 illustrates the results of the images obtained for each method. From left to right, the figures show the original image, the displayed image without image quality improvement, the result of the conventional image quality improvement algorithm (one iteration), and the displayed image with image quality improvement using the proposed method.

Table 1 SSIM results with each method.

algorithm	Image1	Image2	Image3
No image improvement	0.67488	0.59282	0.68393
Conventional method (one iteration)	0.72111	0.62841	0.76684
Proposed method	0.72930	0.71760	0.80679

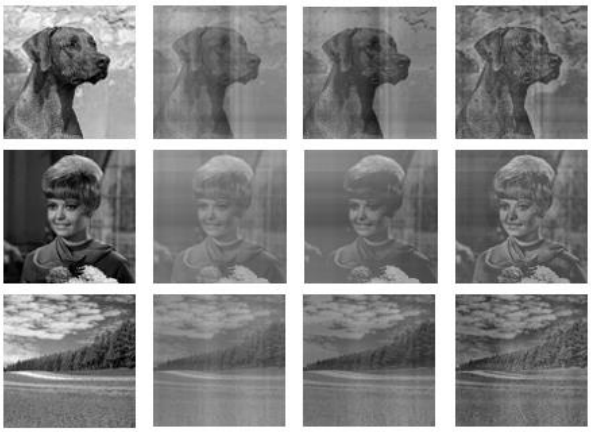


Fig. 7 Results of the images obtained for each method.

4.2 Execution time of Image quality improvement

We measured the time required for their respective processes to evaluate the performance of the conventional image quality improvement algorithm and the image quality improvement algorithm of the proposed method. Conventional image quality improvement algorithms operate by acquiring voxel data and generating output images from input images, with the measurement involving the time taken to acquire new input images using these input and output images. In contrast, the proposed image quality improvement algorithm measures the time required to input an image into the trained model and obtain new image as output. In our experiment, the time required to enhance image quality using the conventional method (one iteration) was approximately 10.68 s. Conversely, the proposed method demonstrated significantly improved efficiency,

with a time of only 0.51 s required to enhance image quality. This remarkable reduction in processing time highlights the improved performance of the proposed image quality improvement algorithm compared to the conventional approach.

4.3 Discussion

When employing the conventional image quality improvement algorithm and the image quality improvement algorithm of the proposed method, noticeable enhancements in the SSIM values were observed. The extent of SSIM value improvement in the proposed method was comparable to that achieved by the conventional algorithm after one iteration. However, the SSIM value is anticipated to increase with each successive iteration of image quality improvement when using the conventional algorithm. The learning model of the proposed method learns the alterations in the pixel values between the original and output images before any quality enhancement. It also improves image quality by using new images obtained from the original images, similar to the conventional method. Because the output of the trained model remains constant concerning the input and disregards other features of the original image, the resultant SSIM value is expected to fluctuate depending on the image combination. By developing a learning model capable of modifying the output image based on a specific combination of original images, we anticipate that image quality can be enhanced to better align with those unique original images.

In terms of execution time, the proposed method takes about 1/20 of the time required by the conventional method. This is because conventional image quality improvement requires that the output image be computed from the original image and then updated by comparing pixel values. Each iteration improves the image quality, but each iteration requires the computation of voxel values and output images. Therefore, the execution time increases in proportion to the number of iterations. In contrast, the proposed algorithm learns a wide range of pixel value relationships between the output and original images, so it does not need to compute voxel values and output images once, which results in short computation time.

5 Conclusions

This research aims to enhance the image quality of directional volumetric displays capable of presenting diverse images in multiple directions using Red-Net. Because of this approach, notable improvements in the SSIM values were achieved. In contrast to conventional image quality improvement algorithms, which require multiple iterations of input image updates by sequentially computing voxel data, the proposed image quality enhancement algorithm generates a new input image by feeding or inputting the original image into a pretrained Red-Net model, resulting in a significant reduction in execution time. Our future objective is to elevate image quality by developing models that consider various combinations of source images and generate diverse outputs tailored to the specific requirements of images intended for use in directional volumetric displays.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 23H03513 and IAAR Research Support Program, Chiba University, Japan.

References

[1] M. Takenaka, T. Kakue, T. Shimobaba, and T. Ito, “Interactive Holographic Display for Real-Time Drawing and Erasing of 3D Point-Cloud Images with a Fingertip”, IEEE Access, 9, pp. 36766-36774 (2021).

[2] S. Eitoku, T. Tanikawa, and Y. Suzuki, “Display Composed of Water Drops for Filling Space with Materialized Virtual Three-Dimensional Objects”, IEEE, pp. 159-166 (2006).

[3] H. Nakayama, A. Shiraki, R. Hirayama, N. Masuda, T. Shimobaba, and T. Ito, “Three-Dimensional Volume Containing Multiple Two-Dimensional Information Patterns”, Scientific Reports, 3, Article number 1931 (2013).

[4] R. Hirayama, H. Nakayama, A. Shiraki, T. Kakue, T. Shimobaba and T. Ito, “Image Quality Improvement for a 3D Structure Exhibiting Multiple 2D Patterns and Its Implementation”, Optics Express, 24, No. 7, pp. 7319-7327 (2016).

[5] A. Shiraki, D. Matsumoto, R. Hirayama, H. Nakayama, T. Kakue, T. Shimobaba, and T. Ito, “Improvement of an Algorithm for Displaying Multiple Images in One Space”, Applied Optics, 58, No. 5, pp. A1-A6 (2019).

[6] X. Mao, C. Shen, and Y. Yang, “Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections”, arXiv:1606.08921 [cs.CV] (2016).