A Deep Learning-Based Approach for Generating 3D Models of Fluid Arts

Hung Mai Cong[1](✉), Mai Xuan Trang[2], Akihiro Yamada[3], Suzuki Takashi[3], Naoko Tosa[4], and Ryohei Nakatsu[4]

[1] Osaka University, Osaka, Japan
hungmcuet@gmail.com
[2] Faculty of Computer Science, Phenikaa University, Hanoi, Vietnam
trang.maixuan@phenikaa-uni.edu.vn
[3] Toppan Inc., Tokyo, Japan
{akihiro_1.yamada,takashi.suzuki}@toppan.co.jp
[4] Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan
{tosa.naoko.5c,nakatsu.ryohei.7r}@kyoto-u.ac.jp

Abstract. This paper explores a method for creating 3D models of video artwork based on a fluid phenomenon called the Sound of Ikebana artwork. A process of using multiple generative adversarial networks (GANs) to reconstruct and predict the shape of the fluid artworks from two-dimensional reference photos was proposed. This is an extension of our previous efforts with Wasserstein GAN enhancements to predict the shape of the unmapped part and correct the texture. The experiment's results show that our process can reconstruct 3D arts without having large amount of 3D training data.

Keywords: 3D fluid art · Sound of Ikebana · DIB-R network · GAN variants

1 Introduction

Recent advances in 3D technology have taken 3D entertainment to a new level. They have also created a new demand for 3D artworks. This paper addresses the issue of creating a 3D model from 2D reference photos of a fluid art.

One of the authors of this paper, Naoko Tosa, has developed an original idea based on fluid phenomena titled "Sound of Ikebana [1]". The Sound of Ikebana is created by capturing Ikebana-like shapes of fluid flows with a high-speed camera that can takes 2000 frames per second. By manipulating the liquid's material and sound, she attempted to convey a wide range of color variations and cultural stories. Despite the fact that the artwork is based on natural phenomena, people can still sense the Japanese beauty in the it. The piece is regarded as one of her most famous well-known creations.

Sound of Ikebana is captured in form of 2D videos and photos. We seek a method to produce the 3D model of the Sound of Ikebana so that people can enjoy the full view

H. Mai Cong and M. X. Trang—Equally contributed as the first author.

A. L. Brooks (Ed.): ArtsIT 2022, LNICST 479, pp. 51–61, 2023.
https://doi.org/10.1007/978-3-031-28993-4_4

of the fluid artwork. However, because the physical limitations in the recording method result from the fast-moving and short-lived properties of the flow, the artist cannot scan the entire 3D scene of the Sound of Ikebana in high quality. Therefore, creating the 3D Sound of Ikebana requires advanced techniques to overcome this limitation.

Our approach uses deep learning models to predict the back and side view of the 2D Sound of Ikebana artwork and build a 3D model based on the predicted information. The pioneering work on this approach is GANverse3D [2], a 3D variant of Generative Adversarial Networks (GANs) [3]. GANverse3D consists of two networks: StyleGAN [4] and an inverse graphics neural network. StyleGAN creates a training dataset based on images of the major objects (persons, cats, dogs, etc.) taken from various angles of an input image; and the inverse graphics neural network deforms a sphere into a predicted 3D model of the input images from the datasets generated by StyleGAN.

The lack of pre-trained 3D information and training data with multiple views of the Sound Ikebana requires us to combine this idea with previous research. The authors in [5] obtained point cloud data for the front aspect of the Sound of Ikebana using Phase Only Correlation approach [6, 7]. In this work, this point cloud data is employed as a reference shape and combine it with several variations of GANs to reconstruct the front view, predict the multi-aspect of the Sound of Ikebana artworks, and create their 3D models.

The paper is organized as follows: A brief introduction of how "Sound of Ikebana" were created is introduced in Sect. 2, Sect. 3 details our improvement with deep learning techniques in creating the 3D Sound of Ikebana. Finally, Sect. 4 discusses the result.

2 Sound of Ikebana

The Sound of Ikebana is a classic illustration of fluid arts. Interestingly, fluid flows have some connection to art. Fluid flows are natural and flexible, and they could represent beautiful forms such as the "milk crown," therefore, it helps artists to create various kinds of shapes. Moreover, the uncertainty of the fluid dynamics gives an artist the enjoyment of unexpected phenomenon that appears in their artworks.

Naoko Tosa used sound vibrations to create fluid flows that gave born of the Sound of Ikebana. The whole system mainly consists of a speaker with corn on top. Then, a thin rubber is put on it, and fluid materials such as color paint is put on the rubber. Then, the sound causes the corn to vibrate, and the liquid jumps up making various shapes. A 2000 frames per seconds (fps) high-speed camera captures the jumping-up liquid phenomenon. The whole system is illustrated in Fig. 1.

Figure 2 shows several typical scenes of the Sound of Ikebana. The artist controls the creation process by changing the sound feature, sound volume, and material of raw paint to create various fluid forms. In addition, the fluid flow's flexibility and uncertainty help the artist create various beautiful forms. The artwork is considered an expression of Japanese art philosophy as one might use various color materials to represent Japanese seasonal flowers and the "Wabi-Sabi" (Japanese sense of beauty which means "beauty within simplicity") aesthetics. For example, Fig. 3 illustrates the similarity between the Sound of Ikebana and the fundamental shape of Ikebana, which is an asymmetrical triangle connecting vertices of various heights: "core", "sub", and "body".

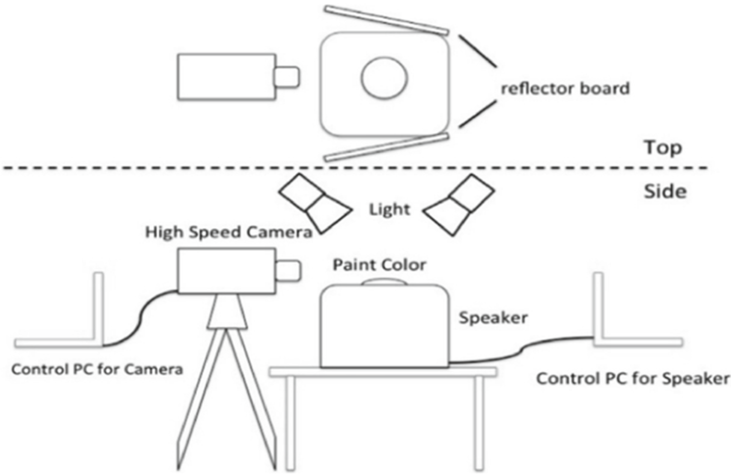

Fig. 1. Sound of Ikebana Generation System [1]

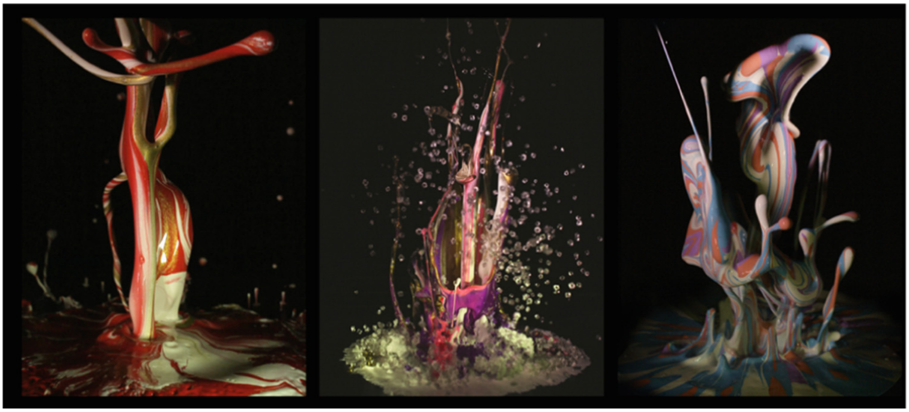

Fig. 2. Sound of Ikebana Artworks

How to extract beauty from natural phenomena has been one big topic in Japanese art. Some notable Japanese works of art were inspired by rivers and the beauty of the scattered waves. The asymmetric triangle is perhaps the most prevalent form of nature, and traditional Ikebana artists undoubtedly strove to reflect nature as simply as possible. In the modern scene of Japanese art, with the help of high technology, Tosa has found that the fundamental form of nature is the same asymmetric triangle. She titled the piece of art "Sound of Ikebana" for this reason.

The fluid flows produced by the sound vibration lasts for a very short time, so the artist had to use a high-speed camera. A 2000 fps camera was used to capture the flows so that it could be reproduced at speed of approximately 67 times slower than actual time. Since the footage was captured with the 2000 fps camera, it is not easy to capture the 3D information of the Sound of Ikebana with today's technology. Therefore, we need

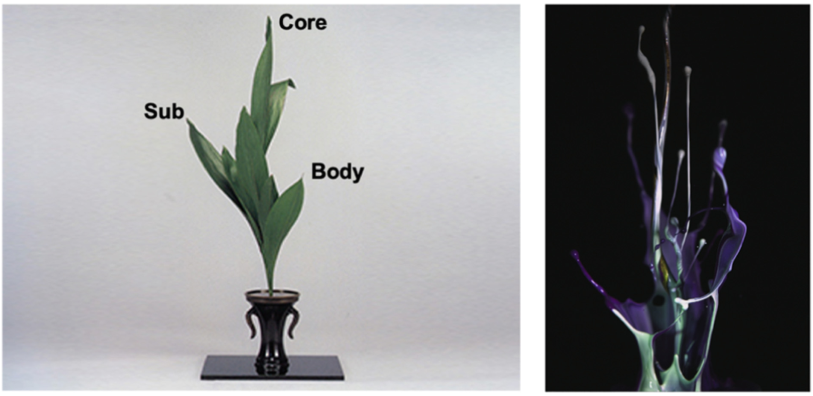

Fig. 3. Left: The Ikebana's basic form. Right: A Sound of Ikebana shape.

to apply some advanced techniques to reconstruct 3D artworks from two-dimensional photos.

3 Generation of 3D Sound of Ikebana Artworks

3.1 Point Cloud Estimation by Phase-Only Correlation Method

The Phase-Only Correlation (POC) approach, which Toppan Printing Inc. has commercialized, is used to estimate the point cloud in the first attempt to produce 3D models of the Sound of Ikebana [5]. The idea is to place several high-speed cameras close to the ink materials' front faces and take several frames of the fluid flow from various angles (as showed in Fig. 4). The point cloud is then estimated from the photo frames using POC method.

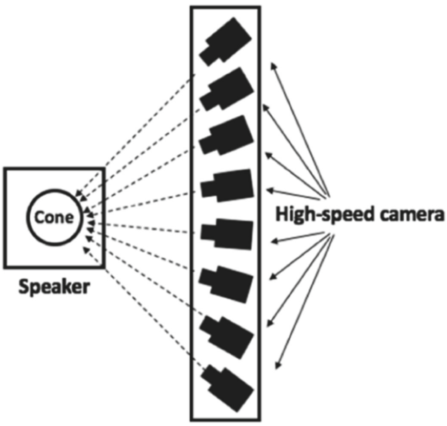

Fig. 4. Multi-camera setting to apply the POC method [5]

In [5], the authors used Poisson reconstruction to generate 3D meshes from the point cloud acquired by POC. Due to the camera angle, these meshes could only accurately represent the front 3D shape of the Sound of Ikebana. The 3D meshes were not ready for 3D printing and required further manual processing to clean up the mesh (see examples in Fig. 5). Therefore, an enhancement to approximate the 3D mesh from the point cloud of the front side of the Sound of Ikebana is essential.

To sharpen the shape of the reconstructed meshes from the point cloud, one can replace the Poisson reconstruction step with the AlphaShape algorithm [8]. An example can be found in Fig. 6, where the shape is very close to the original Sound of Ikebana. As a result, we decided to build the 3D Sound of Ikebana using the front-view surface meshes produced by the AlphaShape algorithm. POC performs well when the camera is set in a narrow baseline. In this case, we could obtain good 3D mesh only for the front view of the point cloud. Therefore, we must improve the texture and predict the shape other than the front view. This improvement is performed using multiple GANs, as described in the next section.

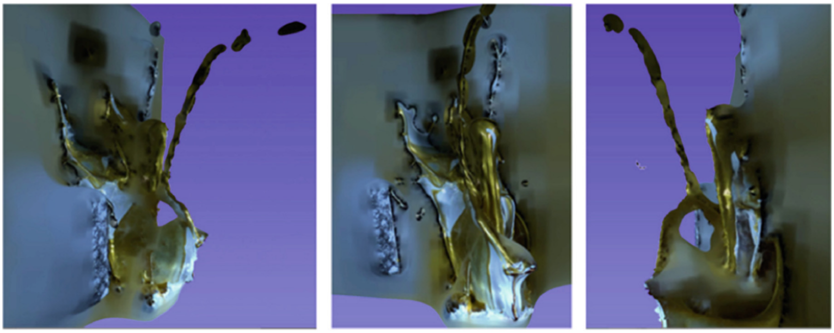

Fig. 5. 3D meshes reconstructed by combining the POC method and Poisson reconstruction [5].

Fig. 6. A surface mesh reconstructed by the AlphaShape algorithm

3.2 Generative Adversarial Networks (GANs)

In recent years, GANs (generative adversarial networks) have become an essential topic in deep learning. The "generative" function of GANs generates new data based on a known data set. A basic GAN network is a combination of two neural networks: A generator network G and a discriminator network D (Fig. 7). In the training of GANs, G tries to generate new data that resembles a target distribution as much as possible. In contrast, D tries to detect whether a data sample is "real" or "fake" as precisely as possible. After this game reaches equilibrium, one might use G to generate new data from random noise or specific input data.

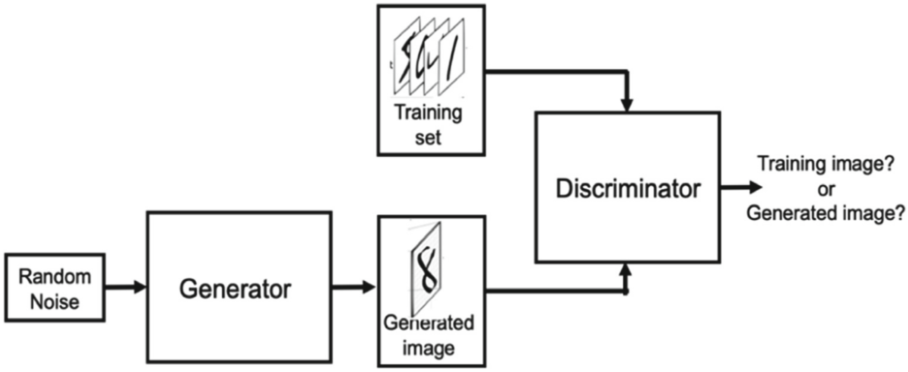

Fig. 7. The primary configuration of GAN ([3])

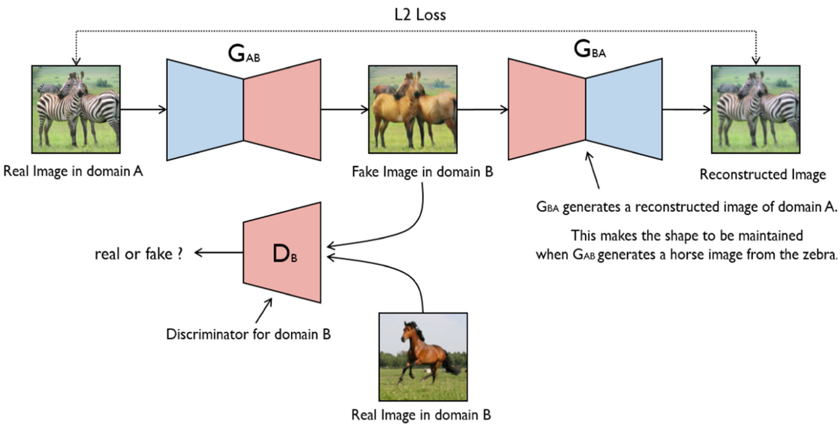

Fig. 8. The CycleGAN's architecture ([10])

One advantage of GANs is that the training process does not require a lot of training data. Numerous GAN variants have been proposed by modifying the basic idea of the minimax game of generators and discriminators to work with different problems. Wasserstein GANs [9] (WGAN) is one of the most used variations of GANs that uses

Wasserstein distance in the loss function instead of cross-entropy in the original GANs. WGAN aids the network avoid mode collapse problems when training GANs.

CycleGAN [10], utilized in style transfer tasks, is another well-known GANs variant. Figure 8 depicts the architecture of CycleGAN, which mutually transforms items between two datasets by an optimized minimax game between generators and discriminators, as well as minimizing the cycle-consistency losses, the identity loss that is difference between an input image and the reconstructed image created by combining two generators. For 3D data, there are several GANs variations (see [11] for example). However, the absence of 3D training data for the Sound of Ikebana makes us consider an approach based on 2D-to-3D GANs variation. We note a pioneer work on this topic - GANverse3D [2]. The concept of this work is to use StyleGAN to generate different two-dimensional photos of an input object from different angles. Then, the system will use these photos as reference information to deform a sphere (via the DIB-R network [12] in the NVIDIA Kaolin library [13]) to obtain an approximated mesh such that the projected images of this mesh are close to the photo generated by StyleGAN. It is challenging to create Sound of Ikebana, obtaining more 2D data from different angles to train StyleGAN is not easy. To deal with this problem we used a combined method that includes the DIB-R network, WGAN, and CycleGAN in our proposed process.

3.3 Proposed 3D Modeling Process

Our proposed 3D generation process for the Sound of Ikebana includes three phases (see Fig. 9). The first phase is to approximate the point cloud data of the front view of the artwork. The second phase involves deforming a sphere into a mesh such that its front view is close to the point cloud of the first phase via the DIB-R network and CycleGAN. The last phase predicts the texture and shape of the unmapped part of the artwork via the DIB-R network and WGAN.

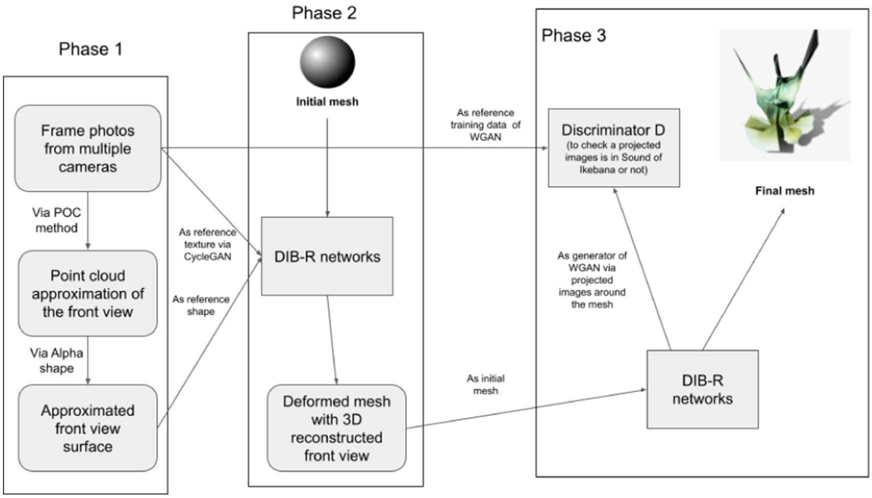

Fig. 9. Proposed 3D modeling process

In our recent research [15], the proposed process only included two phases: the first and the second. In this article, we added the last phase so that a more precise prediction for the side and back view of the Sound of Ikebana would be made and a more precise prediction of their textures.

First Phase

The point cloud for the Sound of Ikebana is initially created using the POC approach. The AlphaShape method is then used to produce a surface mesh replicating the front view of the source artwork.

Second Phase

In the second phase, the front view reconstruction phase (see Fig. 10), the Nvidia Kaolin App is used to generate n 2D projected images ($n = 100$ in our experiment) of the approximated surface mesh from n angles ranging from 0 to 180° in azimuth (the front view) and 0 in elevation. The information about the masks and angles is also stored. Next, we use CycleGAN to transform the Sound of Ikebana's styles into the projected images. Then the DIB-R is used to deform a sphere (the initial mesh) by optimizing the following loss function.

$$L = \lambda_{im} L_{im} + \lambda_{IOU} L_{IOU} + \lambda_{lap} L_{lap} + \lambda_{flat} L_{flat}$$

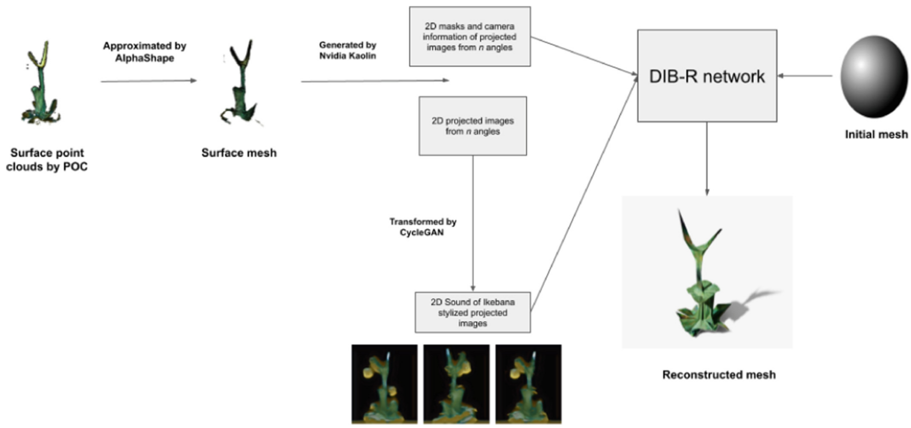

Fig. 10. The front-view reconstruction phase

Here, L_{im} is an $L-1$ loss of the image reconstruction calculated by comparing Ikebana stylized projected images and the projected image of the mesh of the current training stage. L_{IOU} is the intersection-over-union between the ground-truth mask and the rendered mask of the current mesh. L_{lap} and L_{flat} are standard smoothness losses (see [11] for a detailed definition). λ_{im}, λ_{IOU}, λ_{lap}, λ_{flat} are hyperparameters for tuning.

Last Phase

The shape and texture of the uncaptured part of the mesh are predicted using WGAN and DIB-R networks in the final phase. We use Nvidia Kaolin Application to generate randomly numerous viewpoints information (0 in elevation and varies in azimuth). At each training epoch, we obtain projected images of the mesh respected to the angles and update the DIB-R network to continue deforming the mesh in the second phase based on WGAN, where the training data is the frame images captured by multiple cameras. This phase tries to make the projected images look similar to the frame images via WGAN. Here, the DIB-R network plays the role of the generator G in GANs structure, and it generates a new Sound of Ikebana by taking projection images around the mesh concerning the reference angles.

3.4 Results

We obtained the 3D models of some of the Sound of Ikebana's forms (as illustrated in Fig. 12) from the original Sound of Ikebana (as depicted in Fig. 11) by following the process described in Sect. 3.3. This shows that our method successfully reconstructs the front view and predicts the uncaptured parts of the original Sound of Ikebana by referencing the point cloud approximated by the Phase Only Correlation method. These meshes are ready to be printed without additional manual editing. The back-view and side-view are freely transformed but still in harmony with the front view by the transformation based on WGAN.

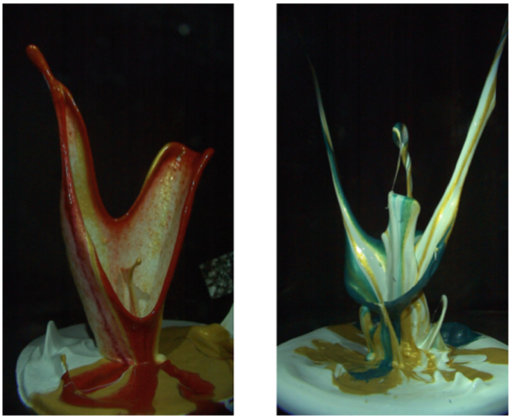

Fig. 11. 2D Sound of Ikebana Images

In our previous attempt [15], we used only phase 1 and phase 2 in the process and reconstructed the mesh based on the point cloud by generating projected images from 0 to 360° (including front, side, and back views). The proposed method can predict the shape of the uncaptured part by comparing it with the shadow of the front view. In this paper, we perform a free transform of the back-view and side-view by adding WGAN to ensure the projected images are in the same style as the Sound of Ikebana by WGAN.

This method helped the final mesh look more natural than the previous work, as the mesh would be asymmetric. The texture is also corrected one more time by WGAN.

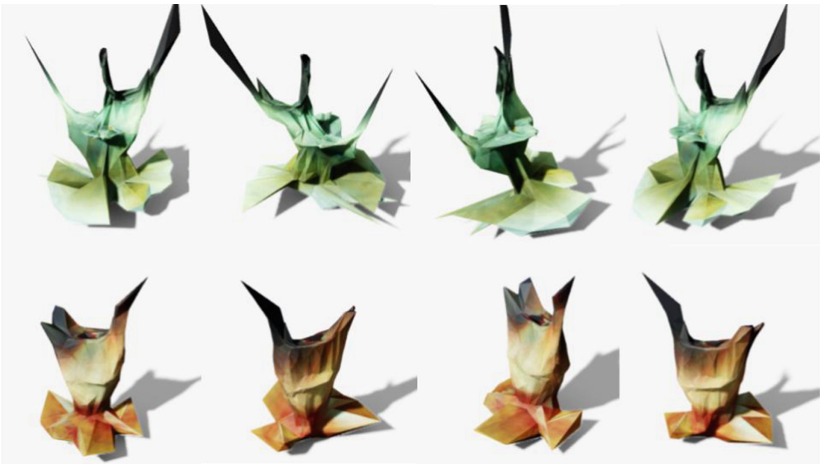

Fig. 12. 3D Sound of Ikebana reconstructed by our process. The left column: the front view. Other columns: images from different angles. (Images captured by the NVIDIA Kaolin App).

Since the DIB-R network is a topological invariant, the 3D Sound of Ikebana has a similar topology to the sphere. In our experiment the initial mesh should be simple, otherwise the deformation may get aggressive or collapse. Therefore, the sphere is chosen as an initial shape. The texture is well transformed, but we expect to generate a smoother texture representing fluid phenomena. Our next experiment will improve the texture quality and expand the work to the Sound of Ikebana with a more complex topology.

4 Conclusion

In this work, we extended the previous efforts in [5] and [15] to build a 3D model of the Sound of Ikebana, a typical example of Fluid Art. The method combines the Phase-Only Correlation method and other deep learning networks such as DIB-R, CycleGAN, and WGAN. Experimental results show that we were able to use multiple deep-learning networks to generate the full 3D Sound of Ikebana without pre-training 3D data. The capability of WGAN helped to improve the prediction of the side view and the back view of the mesh, which is not captured by the high-speed cameras in the creation process of the Sound of Ikebana.

We plan to improve our method for future work to perform a better texture transformation and generate 3D models of the Sound of Ikebana with more complex shapes. Moreover, we expect to generate not only the 3D still model of the Sound of Ikebana but also the 3D videos that represent the moving of color fluid flow in the making of the Sound of Ikebana. The 3D modeling of fluid arts could be applied to various fields of art exhibition, architecture, fashion design, metaverse, etc. Deep Learning technology would be a powerful tool for artists to create arts and industrial designs.

References

1. Tosa, N., Nakatsu, R., Pang, Y.: Creation of media art utilizing fluid dynamics. In: 2017 International Conference on Culture and Computing, pp.129–135 (2017)
2. Zhang, Y., et al.: Image GANs meet differentiable rendering for inverse graphics and interpretable 3D neural rendering. In: ICLR 2021 (2021)
3. Creswell, A., et al.: Generative adversarial networks: an overview. IEEE Sig. Process. Mag. **35**(1), 53–65 (2018)
4. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR 2019 (2019)
5. Tosa, N., et al.: 3D modeling and 3D materialization of fluid art that occurs in very short time. In: 19th IFIP, TC 14 International Conference, pp. 409–421 (2020)
6. Sakai, S., Ito, K., Aoki, T., Masuda, T., Unten, H.: An efficient image matching method for multi-view stereo. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7727, pp. 283–296. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37447-0_22
7. Shuji, S., et al.: Phase-based window matching with geometric correction for multi-view stereo. IEJCE Trans. Inf. Syst. **98**(10), 1818–1828 (2015)
8. Edelsbrunner, H., Kirkpatrick, D.G., Seidel, R.: On the shape of a set of points in the plane. IEEE Trans. Inf. Theory **29**(4), 551–559 (1983)
9. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. https://arxiv.org/abs/1701.07875
10. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251 (2017)
11. Wu, J., et al.: Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In: Advances in Neural Information Processing Systems, pp. 82–90 (2016)
12. Chen, W., et al.: Learning to predict 3D objects with an interpolation-based differentiable renderer. In: 2017 Neural Information Processing Systems NiPS (2019)
13. Fuji Tsang, C., et al.: Kaolin: A PyTorch Library for Accelerating 3D Deep Learning Research (2022). https://github.com/NVIDIAGameWorks/kaolin
14. Tosa, N., et al.: Pursuit and expression of Japanese beauty using technology. Arts **8**(1), 38 (2019)
15. Mai, C.H., Trang, M.X., et al.: Improvement of deep learning technology to create 3D model of fluid art. In: International Conference on Entertainment Computing (ICEC 2022), pp. 227–237 (2022)