Improvement of Deep Learning Technology to Create 3D Model of Fluid Art

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

[1] Osaka University, Osaka, Japan
mai.cong.hung.t3f@osaka-u.ac.jp
[2] Phenikaa University, Hanoi, Vietnam
trang.maixuan@phenikaa-uni.edu.vn
[3] Toppan Inc., Tokyo, Japan
akihiro_1.yamada@toppan.co.jp
[4] Kyoto University, Kyoto, Japan
tosa.naoko.5c@kyoto-u.ac.jp, ryohei.nakatsu@design.kyoto-u.ac.jp

Abstract. Art is an essential part of the entertainment. As 3D entertainment such as 3D games is a trend, it is an exciting topic how to create 3D artworks from 2D artworks. In this work, we investigate the 3D reconstruction problem of the artwork called "Sound of Ikebana," which is created by shooting fluid phenomena using a high-speed camera and can create organic, sophisticated, and complex forms. Firstly, we used the Phase Only Correlation method to capture the artwork's point cloud based on the images captured by multiple high-speed cameras. Then we create a 3D model by a deep learning-based approach from the 2D Sound of Ikebana images. Our result shows that we can apply deep learning techniques to improve the reconstruction of 3D modeling from 2D images with highly complicated forms.

Keywords: Fluid art · Sound of Ikebana · 3D reconstruction · Differentiable rendering network · CycleGAN

1 Introduction

Art has been closely connected to human mentality and has been at the center of entertainment since ancient times. In entertainment such as games, the shift from 2D games to 3D games is a recent trend. Therefore, even in the area of art, how to create 3D artworks from 2D artworks such as paintings is an exciting research theme. We tackle this issue in this paper.

One of the authors, Naoko Tosa, created fluid art based on fluid phenomena. One of her representative fluid artworks, called "Sound of Ikebana [1]," is a collection of video artwork originated by her. The artwork is created by shooting Ikebana-like forms, generated by giving sound vibration to various types of liquid, by a high-speed camera of

M. C. Hung, M. X. Trang---Equally contributed as co-first authors

Published by Springer Nature Switzerland AG 2022
B. Göbl et al. (Eds.): ICEC 2022, LNCS 13477, pp. 227–237, 2022.
https://doi.org/10.1007/978-3-031-20212-4_18

2000 frames/second. She tried to express numerous color variations and cultural stories by utilizing liquid materials.

Although the Sound of Ikebana is a 2D video, obtaining the 3D Sound of Ikebana would give a complete and exciting view of the fluid art. However, the current 3D scanning tools cannot capture the high-speed motion of the fluid flows with high quality.

To reconstruct the 3D model of the Sound of Ikebana, we decided to use the generative models in deep learning. Breakthrough technology in generative models of deep learning has been named Generative Adversarial Networks (GANs) [2]. GANs model includes a generator G that learns to generate new data. At the same time, a discriminator D tries to identify whether the generated data lies in the distribution of the latent space or not. The training process can converge even with a small number of training data by performing learning as a zero-sum game between these two networks. Many GANs variations have been developed by modifying the basic configuration. It is natural to consider applying GANs to 3D generative problems. For instance, one might refer to pioneering works on 3D GANs in constructing 3D models of familiar objects [3].

For artworks like the Sound of Ikebana, the absence of 3D training models requires a different technique. The only training data we have are 2D photos, and we need a 2D-to-3D GAN model instead. For a pioneer work of this approach, we refer to the work of GANverse3D [4]. GANverse3D architecture consists of two neural networks that render images. The first one is StyleGAN [5], which constructs a training dataset based on photos of the main object (vehicles, birds, horses…) taken from different angles to generate multi-view images from an input image. The obtained sets of photos from various angles were fed into an inverse graphics neural network that extracts a 3D model of an object from 2D input images.

In the previous work on the 3D modeling of Sound of Ikebana, Tosa et al. obtained the point cloud data of the front-view surface via the Phase Only Correlation method [6]. In this study, we improve the results via a deep learning approach inspired by the idea of GANverse3D. We perform the deformation from an initial untextured mesh into the targeted 3D Sound of Ikebana using the differentiable rendering network DIB-R [7]. In the process, the 2D Sound of Ikebana stylized images of the back-view and side-view of the mesh generated by CycleGAN [8], a GANs variation with the style transfer capability, is used.

This paper consists of the following sections. In Sect. 2, the concept and creation process of the Sound of Ikebana, a fluid art created by Naoko Tosa, one of the authors. Section 3 details the generation of the 3D Sound of Ikebana, including its process and the obtained result. Moreover, Sect. 4 gives the conclusion.

2 Sound of Ikebana

Sound of Ikebana is a media art creation based on fluid dynamics. Fluid dynamics is a physics discipline that studies the behavior of fluid flows. The fluid flows have some properties that inspire artists to create new art. Firstly, the primary reason for this is that the fluid flows are flexible and natural. Therefore, artists might use them to represent various kinds of shapes. Secondly, fluid dynamics is uncertain even if the initial conditions are fixed. Artists might enjoy the unexpected phenomena or chance phenomena of the fluid flows to incorporate something unexpected into their artworks.

Sound of Ikebana is a typical example of fluid arts created by Naoko Tosa, one of the authors. The artist used sound vibration as the primary method to develop fluid arts. Figure 1 illustrates the Sound of Ikebana generation system. A speaker is set with its corn on top, thin rubber is put on it, and viscous fluid such as color paint is put on the thin rubber. Then, the sound vibrates the corn, and the liquid jumps up, creating various forms. At the same time, the created forms are shot by a high-speed camera with 2000 frames per second.

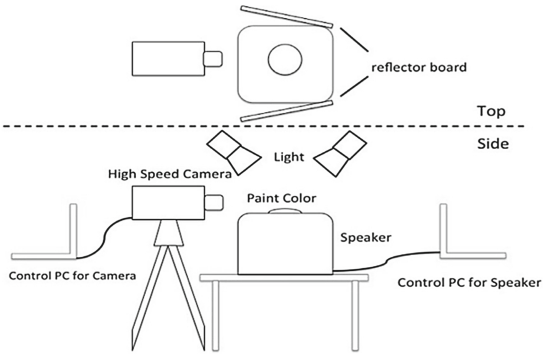

Fig. 1. Sound of Ikebana generation system [1]

While the sound vibration from the speaker controls the initial condition, various color materials make beautiful forms. By carefully choosing the quantity of raw paint and the sound source, the artist successfully created fluid forms that represent the beauty of Japanese Ikebana. The color variations are also considered to express the philosophy of Japanese art, as specific colors are used to represent flowers in each Japanese season and the “Wabi-Sabi” aesthetics in Japanese culture. (Wabi-Sabi is the Japanese sense of beauty, which means “beauty within simplicity.”) Figure 2 shows several images from the artwork.

A high-speed camera is an essential tool in this work. As the fluid flows last only for a significantly short time, the artist used a camera of 2000 frames per second to record the fluid flow. After that, the captured video is replayed at a speed of 30 frames per second, or 67 times slower than the actual time. Then people can enjoy beautiful forms that their naked eyes cannot see. In other words, the recorded videos can visualize hidden beauty in physical or natural phenomena, which we cannot enjoy without high technology.

An interesting issue is why Naoko Tosa named the video art “Ikebana,” a traditional Japanese flower arrangement. She probably found a similarity between Ikebana and her artwork based on her artistic intuition. Figure 3 shows the similarity between the artwork and the primary form of Ikebana. The primary form of Ikebana is an asymmetric triangle that connects three points of different heights, “core,” “sub,” and “body” (Fig. 3: left). It is interesting to note that the shapes in the Sound of Ikebana often resemble the form of Ikebana (Fig. 3: right).

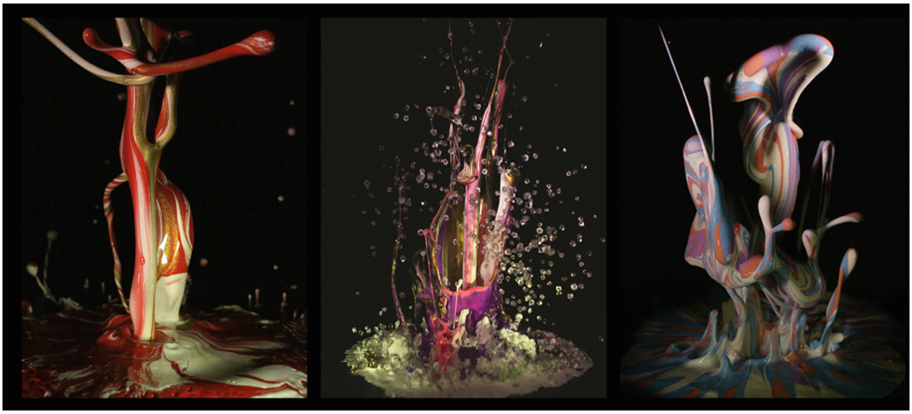

Fig. 2. Examples of the Sound of Ikebana by Naoko Tosa

Japanese artists have found beauty in natural phenomena such as rivers and scattered waves and have created artworks. Ikebana artists in the old days probably tried to represent nature minimally and found that an asymmetric triangle is the primary form of nature. While Tosa attempted to find the primary form of nature by using recent high-speed camera technology, she found the same asymmetric triangle in physical or natural phenomena.

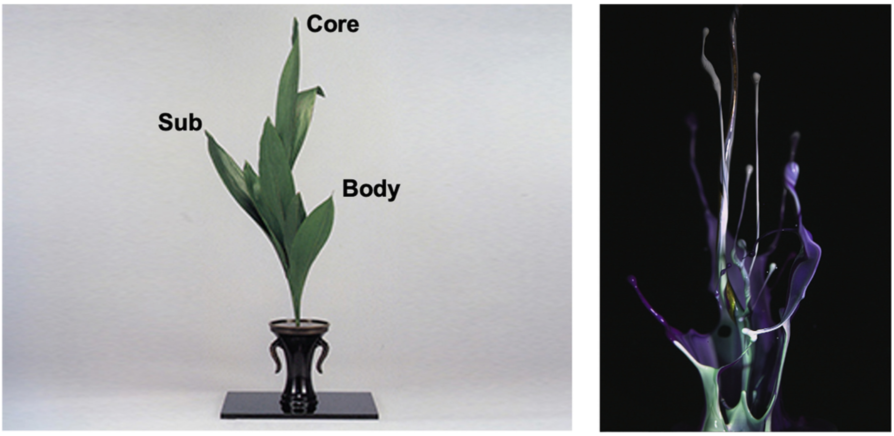

Fig. 3. Left: The basic form of Ikebana (in the public domain). Right: A form of the Sound of Ikebana.

The videos and photos extracted from the Sound of Ikebana are beautiful, but we can only enjoy them from a particular point of view. It is natural to expect to view the complete 3D view of these artworks. If we could do this, we could further investigate the relationship between Ikebana and the Sound of Ikebana. However, the current 3D sensing technologies cannot capture the small-size and fast-moving flows with high quality. Therefore, we need to use some 3D reconstruction techniques from 2D data. In [6], the authors used an image matching technique to extract the point cloud and approximate the 3D mesh from the videos obtained by multiple high-speed cameras. We review this work and improve it. The details are described in the following section.

3 3D Reconstruction of Sound of Ikebana

3.1 Phase only Correlation Method

In the previous work on the 3D modeling of Sound of Ikebana [6], a method using multiple high-speed cameras is introduced. By setting numerous high-speed cameras around the front view of the color materials, one can obtain various photo frames of a fluid flow from multiple viewpoints. Figure 4 illustrates the setting for shooting.

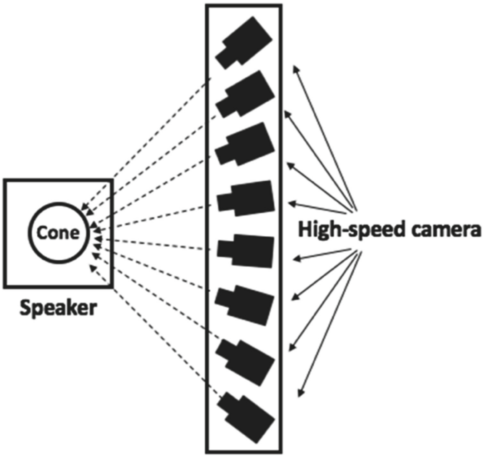

Fig. 4. Multi-camera settings for POC [6]

Then, via the Phase-Only Correlation method [7, 8], these photo frames generate an approximated point cloud of the Sound of Ikebana. Phase-Only Correlation (POC) is an image matching method that uses phase information. The Fourier transform modifies the two images, and a respected POC function is calculated from the transfer spectra. One can obtain an optimal peak model of the POC function when the transformation between two images solely depends on translation. When applying POC to multi-view stereo, k stereo pairs are made from one reference viewpoint and k neighboring viewpoints. POC is utilized to match local windows between each stereo picture. By estimating the coordinates of the 3D points from the peak model of the POC function, the 3D coordinates can be calculated with higher accuracy than conventional methods. This method is effective when the cameras are set in a narrow baseline for complicated shapes like the Sound of Ikebana. Toppan Printing Inc. Has commercialized POC under the name “TORESYS 3DTM”. This software performed the 3D modeling of the Sound of Ikebana in [6].

After obtaining the 3D point clouds of the front view of the Sound of Ikebana via POC, the authors in [6] used the Poisson reconstruction method to approximate meshes from these 3D points. Figure 5 shows the 3D reconstruction obtained from the POC and the Poisson reconstruction method. However, these meshes require further manual editing to obtain a clean 3D mesh, so we sought to improve this method.

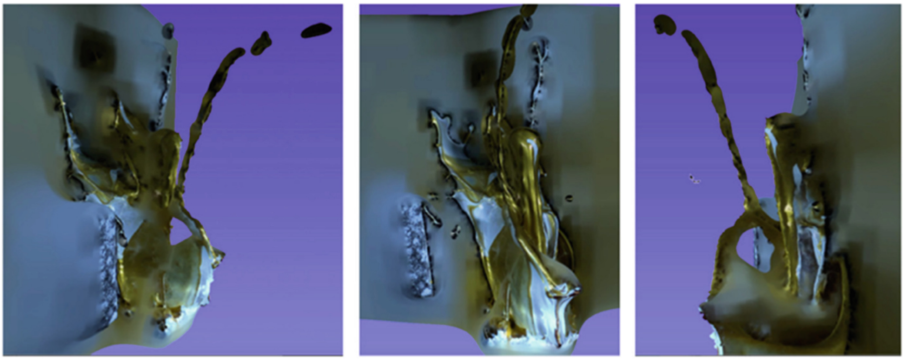

Fig. 5. Examples of 3D Reconstruction by POC and Poisson reconstruction algorithm [6].

When we performed surface reconstruction from the point clouds by the AlphaShape algorithm [9], we obtained surface meshes, as is shown in Fig. 6, which look more similar to the observed Sound of Ikebana. Therefore, we decided to create the 3D Sound of Ikebana models based on the surface meshes obtained by AlphaShape. Since the surface meshes are 2-dimensional manifolds, we consider an approach deforming an initial 3D mesh of a 3-dimensional manifold into an approximated 3D model of the Sound of Ikebana which preserves the fixed surface mesh. We use a deep learning-based method in this task.

Fig. 6. An approximated surface mesh obtained by AlphaShape. Image captured by Meshlab application.

3.2 GANs

GANs (generative adversarial networks) have been an essential topic of deep learning in the last decade. A primary GAN network consists of a generator G and a discriminator D, as shown in Fig. 7. The training process of the network is a minimax game: G tries to generate new data for a training dataset while D tries to evaluate whether a data point is real or fake. The equilibrium state of this game is the point at that G and D could perform their best, and the process does not require extensive training data. Many GANs variations have been developed by modifying the primary mechanism.

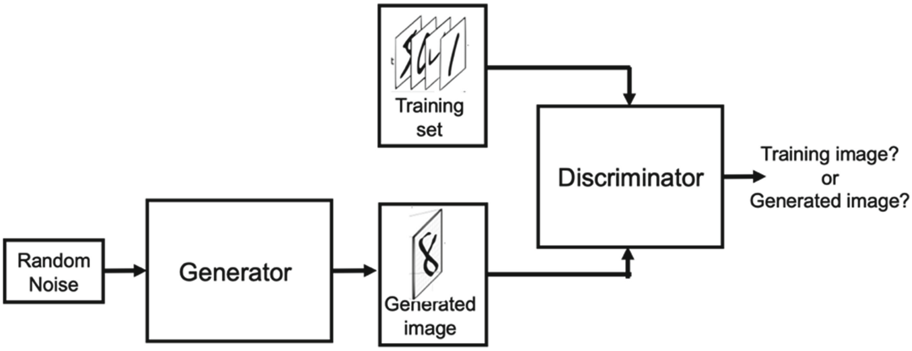

Fig. 7. Basic configuration of GAN ([2])

There are several 3D GANs variations, but we do not have enough 3D data on the Sound of Ikebana at the current level. Therefore, we consider an approach that reconstructs 3D models from 2D images. We refer to the GANverse3D for inspirational work on this task. The authors of GANverse3D use StyleGAN to generate multi-view data from a 2D input image and reconstruct the 3D view by deformation of an initial mesh via the DIB-R network [10], which is included in the Nvidia Kaolin Library [11]. The training process minimizes the loss function consisting of the differences between the multi-view 2D images generated by StyleGAN and the projected 2D images taken around the deforming mesh by the DIB-R network.

Since the Sound of Ikebana is hard to reproduce as fluid flow is uncertain, we could not get enough multi-view data for StyleGAN training. Moreover, we have the approximated 3D surface reconstructed meshes by the POC method and Alpha shape algorithm. So, we could use style transfer to transform the 2D textures of these meshes into the Sound of Ikebana style. Therefore, we use CycleGAN[12] this time. CycleGAN is a set-to-set level transformation network between two data sets as it does not require paired training. The architecture of CycleGAN is illustrated in Fig. 8 [12]. The generator's training process optimizes the minimax game between generators and discriminators. It minimizes the cycle-consistency loss by comparing the difference between an input image and the reconstructed image by combining two generators (from domain A to domain B and domain B).

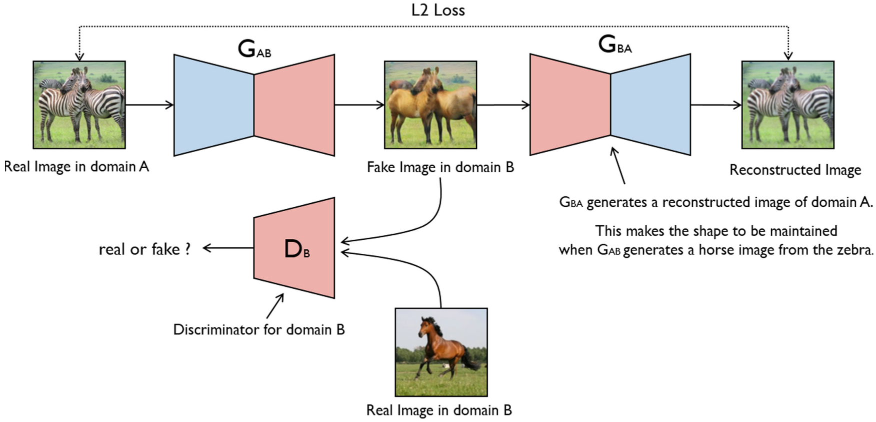

Fig. 8. Basic configuration of CycleGAN ([12])

We use CycleGAN to predict 2D information of the uncaptured flow by transforming 2D projected images taken around the approximated mesh by POC into 2D images of Sound of Ikebana style. The complete process will be discussed in 3.3.

3.3 Improvement of 3D Modeling for Sound of Ikebana

Our experiment includes the following steps, as illustrated in Fig. 9.

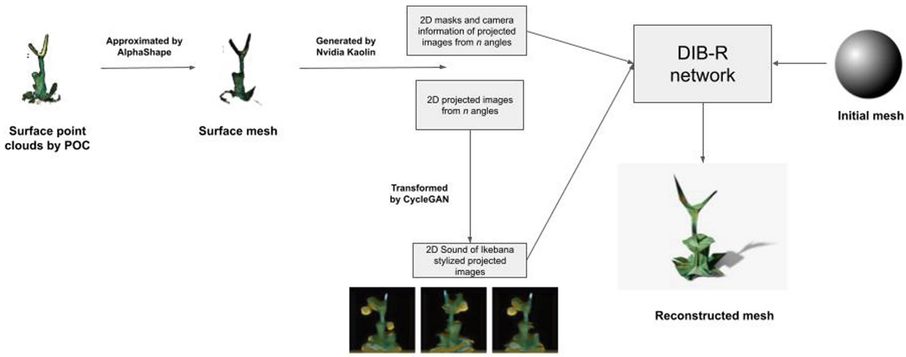

Fig. 9. Our 3D modeling process

Step 1: We use the POC method to obtain the point cloud of the Sound of Ikebana obtained from images with multiple angles.

Step 2: We use the AlphaShape algorithm [9] to reconstruct an approximated surface mesh. We also use Laplacian normalization to smooth out the surface mesh.

Step 3: We use Nvidia Kaolin Application to generate *n* 2D projected images of the approximated surface mesh from *n* angles. Here we fix the elevation to be 0 and vary in azimuth. We also obtain n 2D masks and angle information of these images by the Kaolin Application. In our experiment, we set $n = 200$.

Step 4: We perform CycleGAN transformation between the projected images in Step 3 and the captured frame images of the artwork by multiple cameras to obtain Sound of Ikebana stylized projected images. We use these images to predict the texture of the 3-dimensional manifold mesh.

Step 5: We deform an initial mesh (sphere) via the DIB-R network by minimizing the following loss function:

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

Here, L_{im} is the standard image reconstruction loss between Ikebana stylized projected images in step 4 and the projected image of the current mesh (the projected angles information obtained in step 2) defined in the RGB color space. L_{IOU} is the intersection-over-union between the ground-truth mask (obtained in step 2) and the rendered mask of the current mesh. L_{lap} and L_{flat} are smooth regularization losses (see [11]). λ_{im}, λ_{IOU}, λ_{lap}, λ_{flat} are hyperparameters. We used these hyperparameters for fine-tuning purposes.

3.4 Results

Following the previous subsection process, we obtain the 3D model of some forms of the Sound of Ikebana, as shown in Fig. 11. The original 2D Sound of Ikebana artworks are shown in Fig. 10.

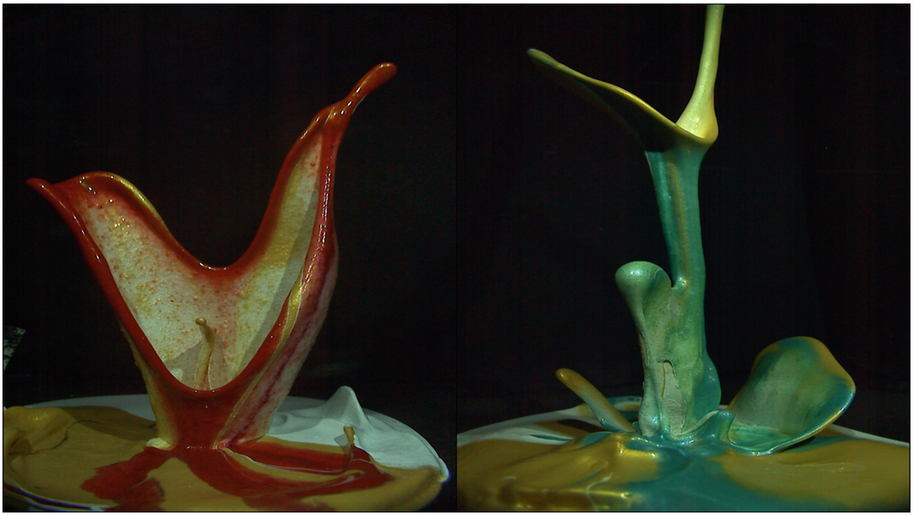

Fig. 10. Original Sound of Ikebana used in our experiment

The results show that our method has reconstructed well-approximated (in terms of shape and texture) 3D meshes from the point clouds obtained by the POC method. These meshes are 3-dimensional manifolds and are ready to print by 3D printers.

Fig. 11. Examples of 3D Sound of Ikebana obtained by our experiment. Image captured by the NVIDIA Kaolin Application.

This experiment reconstructs the Sound of Ikebana artworks with similar topology to the initial mesh (sphere) as the DIB-R network is topology invariant. We note that the initial mesh should not be too complex as the deformation might be aggressive or collapse. We might use different initial meshes (but not too complex) in the following experiments for the Sound of Ikebana artworks of multiple connected components and multiple holes.

The texture of the meshes is well reconstructed at a certain level but not as smooth as the original Sound of Ikebana artworks. Moreover, if we increase the weight of the smooth normalization losses in Step 5, the meshes would not be well deformed into the expected shapes. We consider using an improvision of CycleGAN to increase the resolution of the style transfer step and improve the smooth normalization losses for a better quality of the textures.

As the cameras need to be set in a narrow baseline so that the POC method could perform better, we only obtain the 2D training data of the front view of the artworks. Our method is limited to predicting the meshes' side view and back view in this experiment. Therefore the inner information of the 3D Sound of Ikebana was not reconstructed in this study and requires advanced technique.

4 Conclusion

We have been creating fluid art that utilizes fluid phenomena, led by Naoko Tosa, one of the authors. A representative fluid art called "Sound of Ikebana" is produced by giving sound vibration to fluid such as color paint and shooting the created forms. We found that Ikebana-like beautiful and organic forms are created through this process. It is an exciting research issue to investigate why there is a similarity between the forms of the

Sound of Ikebana and actual Ikebana. For this purpose, it is desirable to create 3D forms of the Sound of Ikebana to observe the forms of the artwork from various viewpoints.

This paper described our attempt to create "3D Sound of Ikebana" from 2D training images. Our method consists of two phases. In the first phase, we use Phase-Correlation Only method to obtain the point clouds. We use improved deep learning to generate meshes based on the point clouds and the 2D Sound of Ikebana method in the second phase. The results show that this method could generate the 3D Sound of Ikebana of simply-topology.

In the subsequent study, we will study the reconstruction of more complex shapes and the improvement in the smoothness of the textures. We might also consider a new setting of the cameras to obtain more inner structure of the 3D Sound of Ikebana artworks.

Our works are still in the early stage, and in the subsequent research, we will study how to improve the quality of the 3D models of more complex structures. In the future, we will continue exploring the application of 3D materialized fluid art to various areas in our society.

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. Creswell, A., et al.: Generative adversarial networks: an overview. IEEE Signal Process. Mag. **35**(1), 53–65 (2018)
3. Wu, J., et al.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. Advances in Neural Information Processing Systems, pp. 82–90 (2016)
4. Zhang, Y., et al.: Image GANs meet Differentiable Rendering for Inverse Graphics and Interpretable 3D Neural Rendering. ICLR 2021 (2021)
5. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. CVPR2019 (2019)
6. 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)
7. Sakai, S. et al.: An efficient image matching method for multi-view stereo. ACCV 2012, LNCS, vol. 7727, pp. 283–296. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37447-0_22
8. Shuji, S., et al.: Phase-based window matching with geometric correction for multi-view stereo. IEJCE Trans. Inf. Syst. **98**(10), 1818–1828 (2015)
9. 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)
10. Chen, W. et al.: Learning to predict 3d objects with an interpolation-based differentiable renderer. In: 2017 Neural Information Processing Systems NiPS (2019)
11. Fuji Tsang, C., et al.: Kaolin: A Pytorch Library for Accelerating 3D Deep Learning Research (2022). https://github.com/NVIDIAGameWorks/kaolin
12. 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)