A Deep Learning Approach to Generate 3D Model of Fluid Art

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

[1]Osaka University, Osaka, Japan
[2]Faculty of Computer Science, Phenikaa University, Hanoi, Vietnam
[3]Toppan Inc., Tokyo, Japan
[4] Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan

[*] Equally contributed as the first author

hungmcuet@gmail.com, trang.maixuan@phenikaa-uni.edu.vn, akihiro_1.yamada@toppan.co.jp, takashi.suzuki@toppan.co.jp, tosa.naoko.5c@kyoto-u.ac.jp, 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. We use multiple generative adversarial networks (GANs) to reconstruct and predict the shape of the fluid artworks from two-dimensional reference photos. This is an extension of our previous efforts with WassersteinGAN enhancements to predict the shape of the unmapped part and correct the texture. The results show that we can apply several deep-learning techniques to create 3D art without 3D training data.
Keywords: fluid art, Sound of Ikebana, 3D modeling, differentiable rendering network, WassersteinGAN, CycleGAN

1 Introduction

Recent advances in 3D technology have taken 3D entertainment to a new level. They have also created a new demand for 3D artwork. This paper addresses the issue of creating a 3D model from 2D reference photos of 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 flow with a high-speed camera that takes 2000 frames per second. By controlling the liquid's material and sound, she sought to express numerous color variations and cultural stories. Although the artwork is based on natural phenomena, people feel the Japanese beauty in the artwork. The artwork is considered one of her most famous artworks.

Sound of Ikebana is a collection of 2D videos and photos. We seek a method to provide the 3D model of the Sound of Ikebana so that people can enjoy the full view 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. Style-GAN creates a training dataset based on photos of the main object (vehicles, birds, horses, etc.) taken from different angles of an input image; and an inverse graphics neural network that deforms a sphere into a predicted 3D model of the input images from the 2D 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 of the front view of the Sound of Ikebana via the Phase Only Correlation method [6][7]. In this work, we use this point cloud data as a reference shape and combine it with several variations of GANs to reconstruct the front view, predict the multi-view of the Sound of Ikebana artworks, and generate their 3D model.

This article is organized as follows: Section 2 introduces the Sound of Ikebana, Section 3 details our improvement in creating the 3D Sound of Ikebana, and Section 4 discusses the results obtained.

2 Sound of Ikebana

As the introduction mentions, the Sound of Ikebana is a typical example 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 uses sound vibration to create fluid flows to create the Sound of Ikebana. The whole system mainly consists of a speaker with corn on top. Then, thin rubber is put on it, and viscous fluid such as color paint is put on the rubber. Then, the sound vibrates the corn, and the liquid jumps up, creating various forms. A 2000 fps high-speed camera captures the jumping-up liquid phenomenon. The whole system is illustrated in Fig. 1.

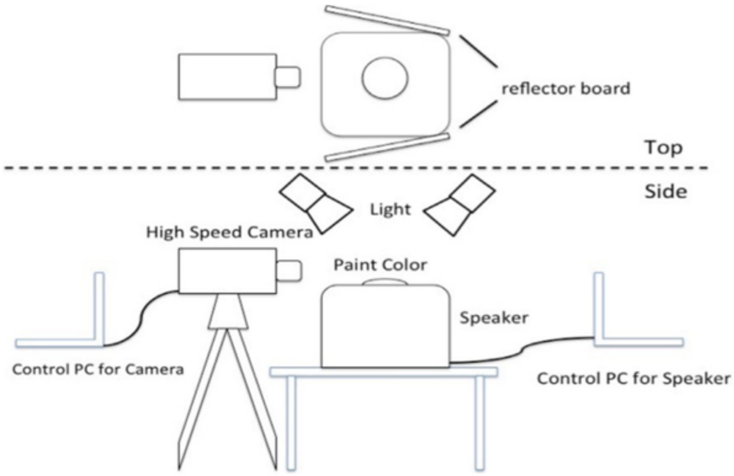


Fig. 1. The system of Sound of Ikebana [1]

Fig. 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" (a Japanese sense of beauty meaning "beauty within simplicity") aesthetics. For example, Fig. 3 shows the similarity between the Sound of Ikebana and the basic form of Ikebana, which is an asymmetrical triangle connecting vertices of different heights: "core," "sub," and "body."

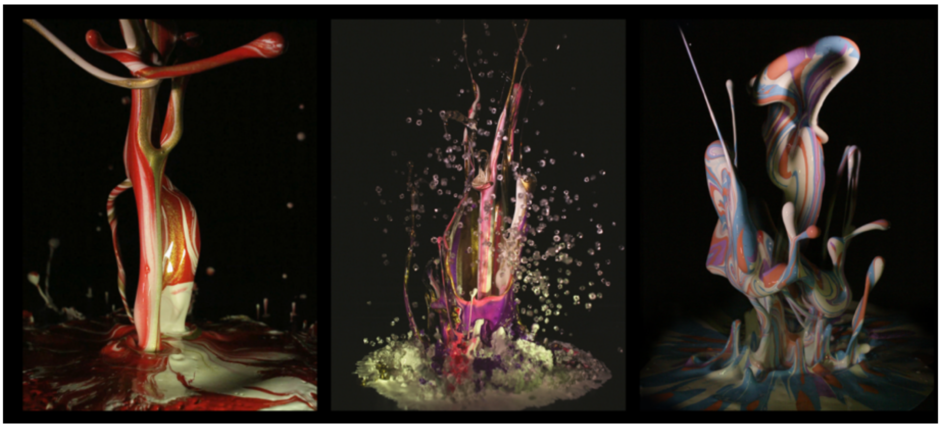


Fig. 2. Some examples of the Sound of Ikebana

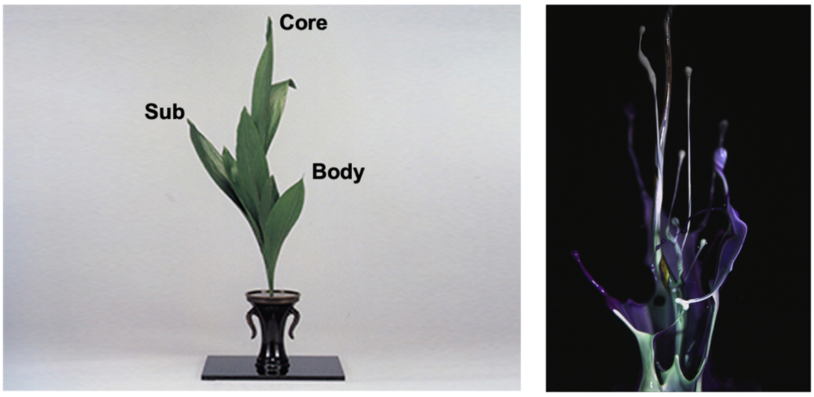


Fig. 3. Left: The basic form of Ikebana. Right: A Sound of Ikebana shape.

How to extract beauty from natural phenomena has been one big topic in Japanese art. Some Japanese artists found beauty in rivers and scattered waves and have created famous artworks based on them. Traditional Ikebana artists probably tried to represent nature minimally and found the asymmetric triangle as the primary form of nature. In the modern scene of Japanese art, with the help of high technology, Tosa has found that the primary form of nature is the same asymmetric triangle. That is why Naoko Tosa named the artwork “Sound of Ikebana.”

The fluid flow produced by the sound vibration lasts for a very short time, so the artist had to use a high-speed camera. Two thousand frames per second camera was used to capture the flow so that it could be reproduced at speed 67 times slower than actual time. Since the footage was captured with 2000 frames per second camera, it is not easy to capture the 3D information of the Sound of Ikebana with today's technology. Therefore, we need to apply some 3D reconstruction techniques from two-dimensional photos. The procedure is described in the following section.

3 Generation of 3D Sound of Ikebana

3.1 Point cloud estimation by Phase-Only Correlation Method

The first attempt to create 3D models of the Sound of Ikebana [5] is to estimate the point cloud using the Phase-Only Correlation (POC) method [6][7], which Toppan Printing Inc has commercialized. The concept is to use multiple high-speed cameras around the front view of the ink materials and capture different frames of the fluid flow from different angles (Fig. 4). Then, the photo frames are used to estimate the point cloud using POC.

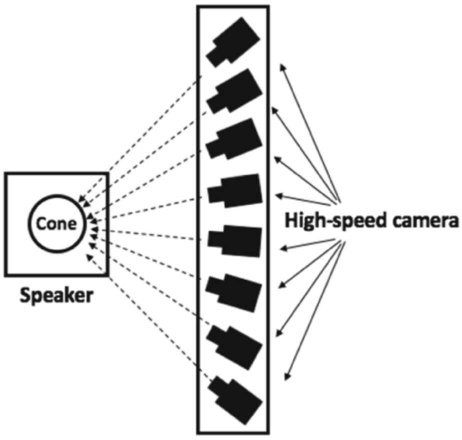


Fig. 4. Multi-camera setting to apply the POC method [5]

In [5], the authors used Poisson reconstruction to generate a 3D mesh from the point cloud obtained by POC. These meshes only reproduced the front 3D shape of the Sound of Ikebana due to the camera setting shown in Fig. 4. Therefore, they 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 point clouds of the front side of the Sound of Ikebana is essential.

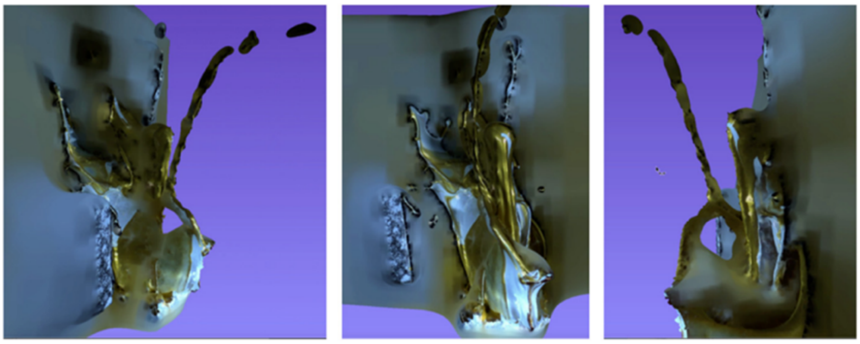


Fig. 5. Some examples of 3D reconstruction by combining the POC method and Poisson reconstruction [5].

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. Therefore, we decided to create the 3D Sound of Ikebana based on the front-view surface meshes obtained by AlphaShape. 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. 6. A surface mesh obtained by AlphaShape captured by Meshlab.

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 GANs network is a combination of two networks: A generator G and a discriminator 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.

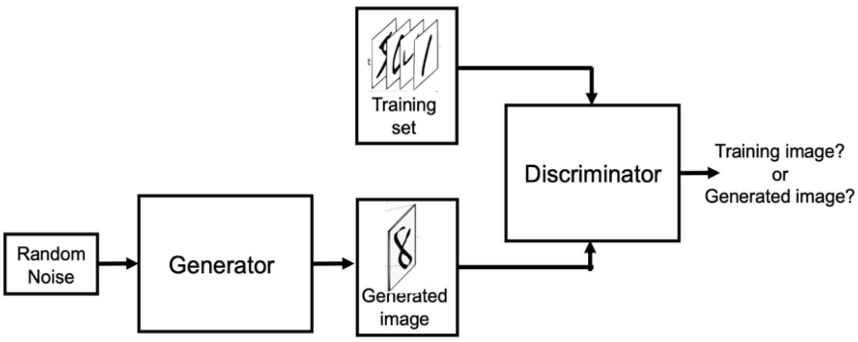


Fig. 7. The primary configuration of GAN ([3])

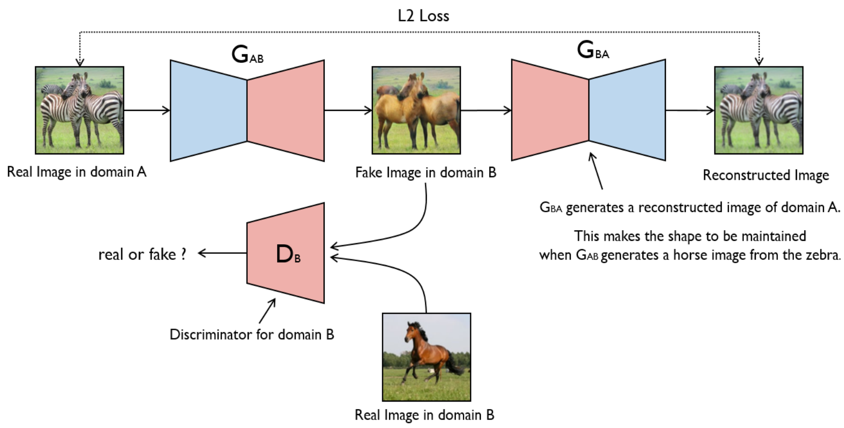


Fig. 8. Basic configuration of CycleGAN ([10])

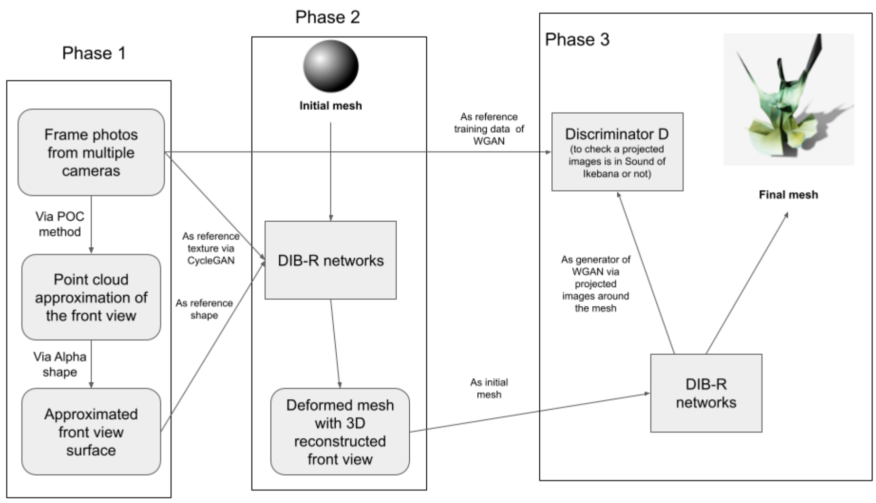
One advantage of GANs is that the training process does not require extensive training data. Many GAN variants have been developed by modifying the basic idea of the minimax game of generators and discriminators to work with different problems. Wasserstein GANs [9] (WGAN) is one of the most used variations of GANs that uses Wasserstein distance in the loss function instead of cross-entropy in the original GANs. WGAN helps the network avoid mode collapse problems when training GANs.

CycleGAN [10], utilized in style transfer tasks, is another well-known GANs variant. Fig. 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 the cycle-consistency loss, the 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. Since the Sound of Ikebana is difficult to reproduce, we needed to obtain more 2D data from different angles to train StyleGAN. Therefore, we used a combined method that includes the DIB -R network, WGAN, and CycleGAN. The detailed procedure is described in the next section.

3.3 Proposed 3D modeling process

Our proposed 3D modeling process for the Sound of Ikebana includes three phases (see Fig. 9). The first phase is to approximate the point cloud information of the front view.

Fig. 9. Proposed 3D modeling process



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.

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

We first utilize the POC method to create the point cloud for the Sound of Ikebana. Next, we recreate a surface mesh representing the front view of the original artwork using the AlphaShape algorithm.

Second phase

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

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

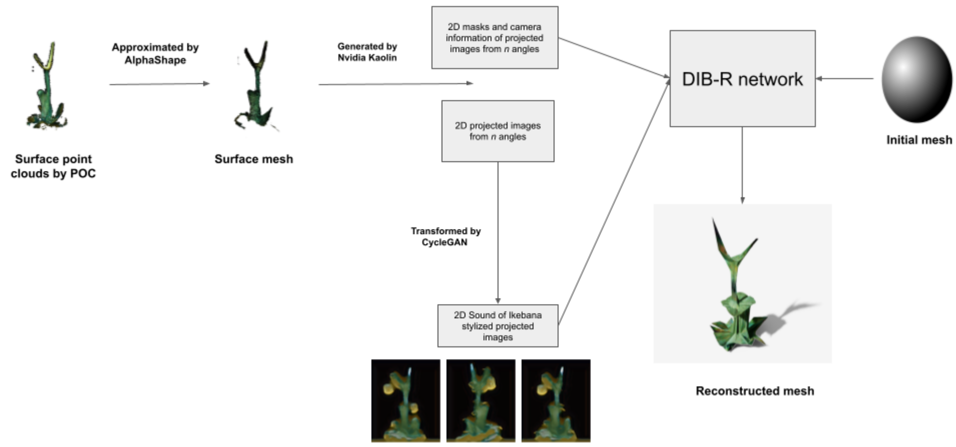


Fig. 10. The front-view reconstruction phase

Here, L_{im} is the standard image reconstruction loss defined 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 smooth regularization 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 model of some forms of the Sound of Ikebana (as shown in Fig. 12) from the original Sound of Ikebana (as shown in Fig. 11) by following the process described in section 3.3. This shows that our method successfully reconstructs the front view and predicts the uncaptured part of the original Sound of Ikebana by referencing the point clouds obtained by POC. 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.

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 degrees (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.

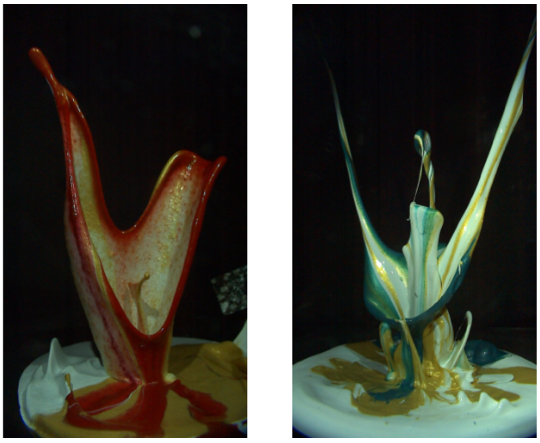


Fig. 11. The original 2D Sound of Ikebana used in the experiment

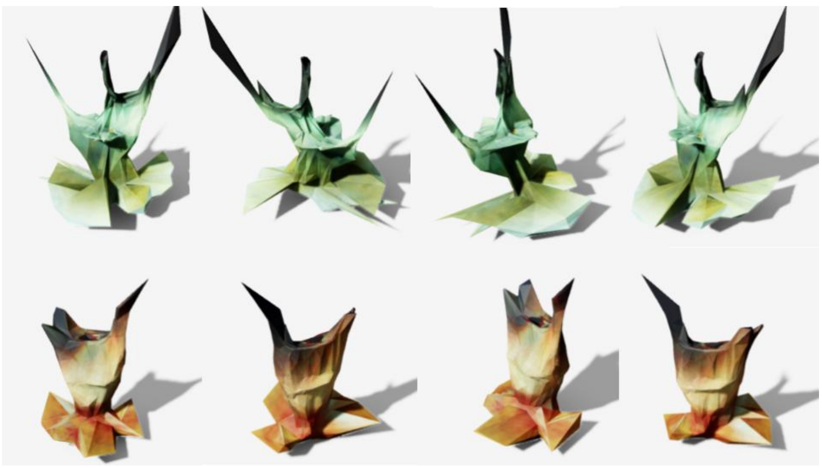


Fig. 12. Examples of the obtained 3D Sound of Ikebana in our experiment. The left column: the front view. Other columns: several images from different angles. Image captured by the NVIDIA Kaolin Application.

As the DIB-R network is a topological invariant, the 3D Sound of Ikebana has a similar topology to the sphere. The initial mesh should be simple once the sphere is chosen as an initial shape. Otherwise, the deformation might be aggressive or collapse. 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 create a 3D model of the Sound of Ikebana with 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 and fashion design, metaverse, etc. Deep Learning would be a powerful tool for artists to create art and industrial design.

References

1. N. Tosa, R. Nakatsu, and Y. Pang, “Creation of Media Art Utilizing Fluid Dynamics,” 2017 International Conference on Culture and Computing, pp.129-135 (2017).
2. Y. Zhang et al., “Image GANs meet Differentiable Rendering for Inverse Graphics and Interpretable 3D Neural Rendering,” ICLR 2021 (2021).
3. A. Creswell et al., "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, Vol.35, No.1, pp.53–65 (2018).
4. T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” CVPR2019 (2019).
5. N. Tosa et al., “3D Modeling and 3D Materialization of Fluid Art that Occurs in Very Short Time,” 19th IFIP, TC 14 International Conference, pp. 409–421 (2020).
6. S. Sakai 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.
7. S. Shuji et al., “Phase-based window matching with geometric correction for multi-view stereo,” IEJCE Trans. Inf. Syst. 98(10), 1818–1828 (2015).
8. H. Edelsbrunner, D. G. Kirkpatrick and R. Seidel, “On the shape of a set of points in the plane,” IEEE Transactions on Information Theory, 29 (4), pp 551–559 (1983).
9. M. Arjovsky, S. Chintala and L. Bottou, “Wasserstein GAN”, https://arxiv.org/abs/1701.07875

10. J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251 (2017).
11. J. Wu 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).
12. Chen, W. et al., "Learning to predict 3d objects with an interpolation-based differentiable renderer," 2017 Neural Information Processing Systems NiPS (2019).
13. C. Fuji Tsang et al., "Kaolin: A Pytorch Library for Accelerating 3D Deep Learning Research," https://github.com/NVIDIAGameWorks/kaolin (2022).
14. N. Tosa, et al., "Pursuit and Expression of Japanese Beauty Using Technology." Arts, Vol.8 No.1, 38 (2019).
15. C.H. Mai, M.X. Trang, et al., "Improvement of Deep Learning Technology to Create 3D Model of Fluid Art," International Conference on Entertainment Computing (ICEC2022) pp 227–237 (2022).