Motion Reproduction of Sky and Water Surface From an Omnidirectional Still Image

Motoki Kakuho, Hakim Ikebayashi and Norihiko Kawai
Graduate School of Information Science and Technology
Osaka Institute of Technology
Hirakata, Osaka, Japan
m1m23a07@st.oit.ac.jp, norihiko.kawai@oit.ac.jp

***Abstract*—Applications such as Google Street View allow users to experience the atmosphere of a place without physically being there. However, one issue with the applications is that the displayed images are still images, which lack a sense of presence when users see them in VR. In this study, we propose a method to enhance the sense of presence by focusing on sky and water areas in an omnidirectional still image and reproducing their motion for video generation.**

***Index Terms*—Omnidirectional Image, Video Generation, Motion Reproduction**

I. Introduction

Applications using omnidirectional images such as Google Street View allow users to view the scenery of a location without actually going there. However, such applications present static images, lacking a sense of presence. Although one of the solutions to the problem is to capture a video from a fixed point, it requires a significant amount of time to capture scenes of all over the world. In this study, we propose a method to generate an omnidirectional video in which the motion of sky and water is reproduced.

Conventional methods for generating videos from a still image can be classified into two types: those that focus on non-fluid objects such as cars and people [1] and those that focus on fluid objects such as fire and water [2], [3]. This study focuses on the latter methods. Among the methods, Endo et al., for example, use neural networks to reproduce the motion of the sky and rivers from a single landscape image. Since the method uses perspective projection images as training data, the resulting image obtained by applying it to an omnidirectional image may contain unnatural motion. In addition, the method is highly dependent on parameters and may generate motion even in areas that are originally stationary.

We reproduce natural motion in an omnidirectional image by the combination of optical flow calculation, which considers the motion in a 3D space rather than neural networks, and inpainting, which estimates sky and water textures of the entire hemisphere. In addition, we use semantic segmentation [4] to clearly distinguish between moving and stationary areas.

II. Proposed method

The flow of the proposed method is as follows. (1) We first input an omnidirectional still image, and (2) apply semantic segmentation to it to create the mask image. (3) We then create the texture of the entire sky by inpainting. (4) We compute the motion of sky and water surface, and create an image sequence frame by frame from the inpainted image. (5) We finally generate a video in which only the sky and the water surface move by combining the image sequence with the input image. We desrcibe the details in the following.

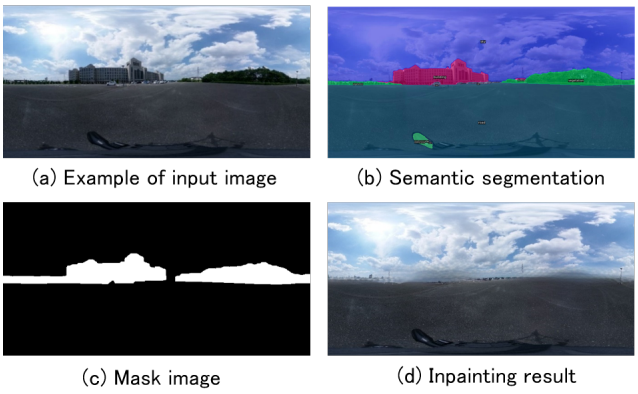

Fig. 1. Example of input image.

First, (1) we input an omnidirectional landscape image containing either sky or water, as shown in Fig. 1(a). This study assumes that omnidirectional images are generated by equirectangular projection so that the bottom pixel is in the direction of gravity obtained from the accelerometer in the camera. Next, (2) we apply the semantic segmentation [4] to the image to divide it into regions such as sky, water surface, and others as shown in Fig. 1(b). From the segmented image, we generate a mask image that mask all objects above the horizon except the sky area, as shown in Fig. 1(c). Here, due to inaccuracies near the boundaries of the semantic segmentation, the mask regions are expanded to fully include objects except the sky area. Next, (3) using the generated mask image, we generate an image in which all areas above the horizon have sky textures by inpainting [5], as shown in Fig. 1(d).

Next, (4) we calculate the motion of sky and water surface. For the sky motion, we use the assumption that the clouds in the sky move straight on the plane above in 3D space, and represent their 3D motion as a 2D optical flow on the omnidirectional image. Specifically, as shown in Fig. 2, the relationship between a pixel (p_1,q_1) in the omnidirectional

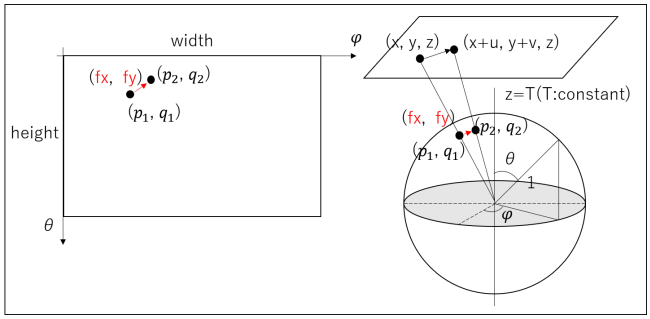

Fig. 2. 3D positional relationship between the sphere and the sky plane.

image and its corresponding position (x, y, z) on the plane above is calculated from the line passing through the center of the sphere and the intersection of the sphere and the plane. Next, a certain distance u and v are added to the x and y coordinates on the plane. By projecting the coordinates $(x + u, y + v, z)$ onto the sphere, the corresponding pixel (p_2,q_2) is calculated. The difference between pixel (p_2, q_2) and pixel (p_1, q_1) is represented as the optical flow (f_x, f_y). By copying the pixel values of the input image using the optical flow obtained while increasing u and v linearly, we generate a frame sequence in which the sky moves. For the water surface, the motion is also generated in the same way by preparing a plane representing the water surface under the sphere.

(5) We finally combine the video generated in (4) with the input image using the mask image as shown in Fig. 1(c) to generate a video in which only the sky and water surface move. Here, alpha blending is performed at the boundary of the mask to reduce the unnaturalness at the boundary between the moving and static regions.

III. Experiments

We conducted experiments using two scenes: Scene A as shown in Fig. 1 and Scene B with both sky and water surface regions. We captured omnidirectional images with an omnidirectional camera (RICOH THETA Z1) and resized them to 1600 × 800 pixels. The results were compared with those obtained by the conventional method [2].

Figure 3 shows the resulting omnidirectional frames and their perspective projection images in a certain direction. From the results, we can see that the sky region moves in the lower right direction. The overhead regions of the sky move faster than the more distant regions, producing natural motion.

Figure 4 shows the results in perspective projection view by the conventional [2] and proposed methods for Scene B. In the results of the conventional method shown in the upper part of Fig. 4, when the direction of the left and right ends of the omnidirectional image is viewed as a perspective projection image, the boundary is clearly recognizable and unnatural because the conventional method generates different movements at the both ends. On the other hand, the proposed method uses 3D information to generate motion, resulting in consistent motion and no obvious boundaries, as shown in the lower part of the Fig. 4. In this scene, we also confirmed

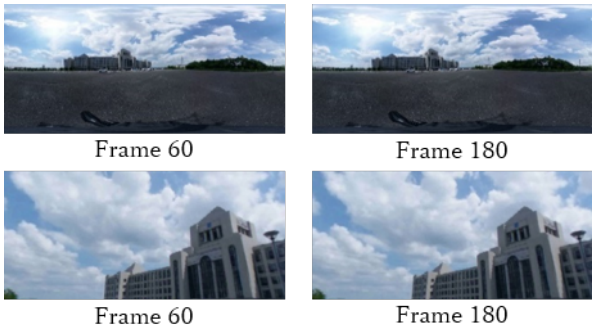

Fig. 3. Results for Scene A.

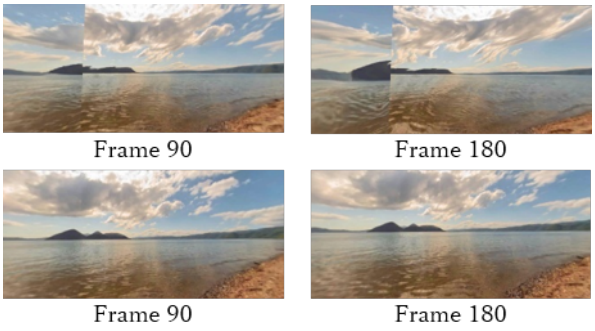

Fig. 4. Results and comparison for Scene B.

that the water naturally moved toward the coastline. However, the waves of the water are not fully represented, resulting in a sense of discomfort. Therefore, to generate more natural water surface motion, it is necessary to consider the 3D motion of waves and the complex motion at the edge of water.

IV. Conclusion

In this study, we proposed a method for reproducing motion of sky and water surface from an omnidirectional still image. Experiments demonstrated the effectiveness of the proposed method. Future work includes making the motion of water more realistic and adding motion to other objects such as trees to create more realistic video.

Acknowledgment

This research was partially supported by JSPS KAKENHI JP21H03483.

References

[1] M. Babaeizadeh, C. Finn, D. Erhan, R. H, Campbell, and S. Levine, "Stochastic variational video prediction," in *ICLR*, 2018.
[2] Y. Endo, Y. Kanamori, and S. Kuriyama, "Animating landscape: Self-supervised learning of decoupled motion and appearance for single-image video video synthesis," *ACM Trans. on Graphics*, vol. 38, no. 6, 2019.
[3] M. Okabe, K. Anjyor, and R. Onai, "Creating fluid animation from a single image using video database," *Computer Graphics Forum*, vol. 30, no. 7, pp. 1973–1982, 2011.
[4] J. Lambert, Z. Lie, O. Sener, J. Hays, and V. Koltun, "MSeg: A composite dataset for multi-domain semantic segmentation," in *CVPR*, 2020.
[5] J. Y. ans Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in *ICCV*, 2019.