

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

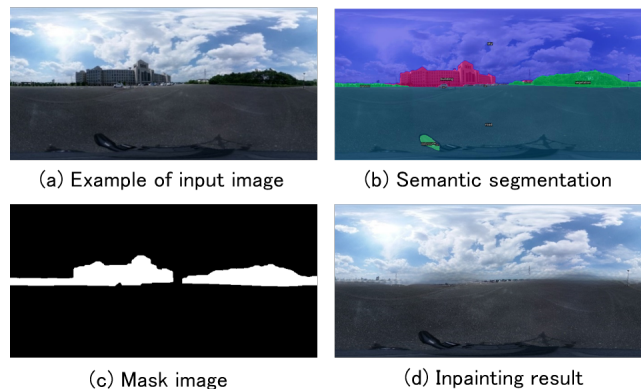


Fig. 1. Example of input image.

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 describe the details in the following.

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

