Real-time Generation of Novel Views of a Dynamic Scene Using Morphing and Visual Hull

Tomoya ISHIKAWA, Kazumasa YAMAZAWA, and Naokazu YOKOYA

Graduate School of Information Science,
Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0192, Japan,
{tomoya-i, yamazawa, yokoya}@is.naist.jp

Abstract—Recently, generation of novel views from images acquired by multiple cameras has been investigated. It can be applied to telepresence effectively. Most conventional methods need some assumptions about the scene such as a static scene and limited positions of objects. In this paper, we propose a new method for generating novel view images of a dynamic scene with a wide view, which does not depend on the scene. The images acquired from omni-directional cameras are first divided into static regions and dynamic regions. The novel view images are then generated by applying a morphing technique to static regions and by computing visual hulls for dynamic regions in real-time. In experiments, we show that a prototype system can generate novel view images in real-time from live video streams.

Keywords-Novel View Generation; Morphing; Visual Hull; Omni-directional Camera

I. INTRODUCTION

Recently, generation of novel view images from images acquired by multiple cameras has been widely investigated [1,2,3,4]. The technique enables telepresence which presents a virtualized real scene to users staying in a remote site to provide a feeling of walking in the remote scene. In the previous work [5,6], we have proposed a telepresence system using omni-directional images, which enables a user to see a virtualized dynamic scene in arbitrary direction. The system acquires omni-directional images from a catadioptric omni-directional camera or an omni-directional multi-camera system in a remote site. The viewpoint is restricted at the position of the omni-directional camera. In the present study, we aim at developing an extended system which enables a user to see a virtualized remote site at arbitrary viewpoints interactively.

In related works, Saito et al. [1] have proposed a method which generates a novel view image in the 3D room. The 3D room mounts a large number of cameras on the walls and ceilings. In the 3D room, they can acquire multiple images of dynamic objects and generate novel views from the images. Saito's method is based on the assumption that the positions of objects are limited within the area covered by the cameras. Inamoto [2] and Koyama [3] have proposed methods for novel view image generation in a soccer scene. Inamoto's method [2] uses characteristics of soccer scenes. In the soccer scene, the soccer field, the far background, and the soccer goal are considered as a single plane or a set of planes. The method can generate high-quality novel view images based on such characteristics of the soccer scene. Therefore, it is difficult to

apply the method to other scenes. Koyama's method [3] is based on the assumption that a viewpoint is far from soccer players. The method approximately generates a novel view image using the billboarding technique in online process. This method is also difficult to be applied to other scenes. Tomite [4] has reported a method for generating an omni-directional novel view image using a morphing technique and omni-directional images acquired by omni-directional cameras. The generated image is omni-directional and thus a user can look around the scene in arbitrary direction. The method requires point correspondences among input images in advance. Therefore, it is difficult to apply the method to a dynamic scene.

In this paper, we propose a new method for generating novel view images of a dynamic environment with a wide view, which does not depend on the scene. The images acquired from omni-directional cameras are first divided into static and dynamic regions. The novel view images are then generated by applying a morphing technique to static regions and by computing visual hulls for dynamic regions in real-time. In experiments with a prototype system, we show that the proposed method can successfully generate novel view images in real-time from live video streams.

Section II introduces the algorithm of the proposed method. Section III gives a detail of experimental environment and shows some examples of generated images. Conclusions and future works are finally given in Section IV.

II. GENERATION OF NOVEL VIEW IMAGE

A. Overview

Fig.1 illustrates an overview of the proposed method. Our approach is as follows. Firstly, omni-directional images are acquired by omni-directional cameras which are located in an environment for capturing a wide area. Secondly, the acquired images are segmented into static and dynamic regions. In the segmentation process, our method uses the robust background subtraction [7] for detecting dynamic objects for a long time. It considers varying lighting conditions and renews a background intensity model in every input frame. Thirdly, after segmentation process, our method generates novel view images of each region by using feasible method for each region. It is necessary to use techniques with geometrical accuracy for generating a novel view image which represents correct

positions between static and dynamic objects when synthesizing each novel view image. The techniques with geometrical accuracy enable to construct binocular stereo by generating images of viewpoints of both eyes. For generating a static novel view image, we uses a morphing technique [4] based on 3-D positions of corresponding points. It enables to generate a novel view image using GPU functions without high CPU cost. In addition, point correspondences for morphing are given in offline process. On the other hand, dynamic regions require a different process for generating a novel view image in every input frame. We use a visual hull based technique for generating a novel view image of dynamic regions. Visual hull is often constructed by using the shape-from-silhouette technique with a voxel space. In the case of computing a wide area, the technique needs huge data and CPU cost. Therefore, our method employs the Image-Based Visual Hull technique [9] without a voxel space. In addition, the novel view image is generated effectively by using positional presumption of dynamic objects simultaneously. Finally, the static novel view image as a background and the dynamic novel view image as a foreground merged into a complete novel view image. In this study, it is assumed that omni-directional cameras are fixed and their position and posture are known. We explain the details of the four processes consisting of segmentation of images, morphing of static regions, computing visual hulls, and synthesis of novel view images in the following.

B. Segmentation of Images into Static and Dynamic Regions

The proposed method divides input images into static and dynamic regions, and generates a novel view image of each region. A simple strategy for division into static and dynamic regions is based on background subtraction which uses the background captured in advance. This strategy suffers from changes of lighting effect and other noises. Our method uses the robust background subtraction [7] which can compensate a flicker of a fluorescent light and CRT and a noise according to gain-up of a camera [7].

This method preserves the average and variance of intensity of the background pixels at each pixel of input images in time series. If the intensity of pixel is within the range of a given background intensity model, the pixel is in a background static region, otherwise, the pixel is in a dynamic region. If a pixel judged to be in a dynamic region represents a static object added into the environment, the pixel is automatically fused to a static region after a while. We model the intensity of background pixels as follows:

$$I = \bar{I} + \sigma \sin(2\pi w t) + k\zeta, \qquad (1)$$

where

I : intensity of background pixel,
$\bar{I}$: average of I in time,
σ : amplitude of oscillation of intensity,
ω : frequency of oscillation of intensity,
t : time,
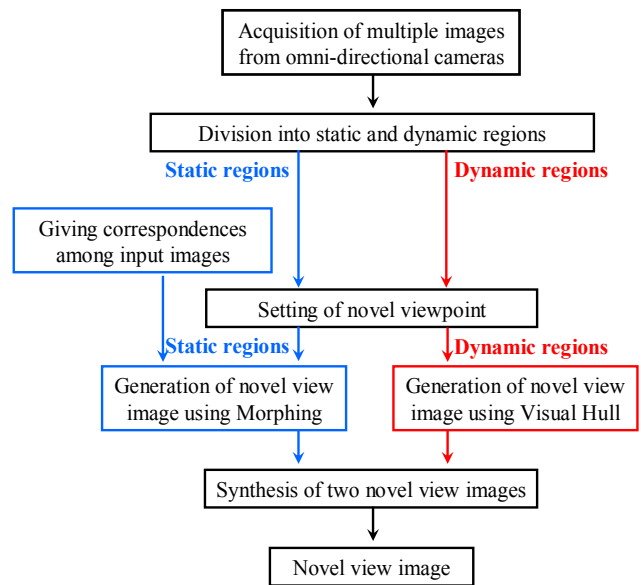k : -1, 0, or +1,
ζ : noise dependent only on the camera.

Fig.1. Overvieww of the proposed method.

The term $\sigma \sin(2\pi w t)$ in Eq.(1) means a flicker of a fluorescent light, CRT, etc. The term $k\zeta$ means a noise, which is dependent only on the camera. Therefore, the background intensity has the range of

$$\bar{I} - \sigma - \zeta \le I \le \bar{I} + \sigma + \zeta. \qquad (2)$$

If the intensity I of the pixel is outside the range given by Eq.(2), we decide that the pixel is a part of a dynamic region. This method renews the parameters I and σ in every frame of an input image sequence. We predetermine the camera-dependent parameter ζ.

C. Novel view generation for static regions

The proposed method uses Tomite's method [4] which generates an omni-directional novel view image of static regions using a morphing technique from omni-directional images. The method requires point correspondences among input images. Note that the positions and postures of multiple omni-directional cameras are known and point correspondences are given among omni-directional images in advance. The process of the image generation is as follows.

Step1. The 3D positions of the corresponding points are computed by omni-directional stereo.
Step2. The 3D points computed by Step1 are projected onto the novel view image plane. (see Fig.2)
Step3. Triangulated patches in the omni-directional novel view image plane are generated based on the projected points using Delaunay's triangulation.
Step4. The omni-directional novel view image is generated by blending the parts of input omni-directional images which correspond to the generated triangles.

In this method, the processing for pixels within triangles and the blending are executed by using OpenGL functions with GPU. Therefore, this method can generate a novel view image without high CPU cost.

D. Novel view generation for dynamic regions

The novel view image of dynamic regions is generated by computing visual hulls. Visual hull constructed by the intersection of silhouette cones from several viewpoint images defines an approximate geometric representation of an object. In general, the shape-from-silhouette technique is used for computing the visual hull [8]. The method employs the voxel representation of space. Therefore, the cost of computing the visual hull and the size of data becomes huge in the case of computing a wide area. The proposed method uses the Image-based Visual Hull technique [9] for computing the visual hull. The technique does not use voxels but generates a novel view image by estimating the penetration of visual hull by the ray from the optical center of virtual camera. The overview of the process is as follows.

Step1. For each pixel in the novel view image, the ray connecting the pixel and the optical center of virtual camera is projected onto input images.

Step2. On the line of ray, we search a region in which all of the projected lines intersect with dynamic regions. If such a region is found, the ray is judged to penetrate the visual hull. If such a region is not found, the ray does not penetrate the visual hull (see Fig.3).

Step3. If the ray penetrates the visual hull, the pixel in novel view image is colored. The point on the intersection region nearest to the novel viewpoint is projected onto the input image that is selected based on the similarity of the angle formed by the point, the novel viewpoint, and the viewpoint of input image. The color of the projected point is decided as the color of the pixel of the novel view.

Step4. The process of Step1-3 is executed for all the pixels in the novel view image.

We can generate the novel view image of dynamic regions. However, when all the pixels in the novel view image are estimated by this process, the computational cost is large. For reducing the computational cost, we estimate the positions of dynamic objects in the environment and calculate an approximate region for generating a novel view image. We use Morita's method for estimating the positions of dynamic objects [7].

E. Synthesis of novel view images

The morphing and computing visual hull process generates novel view images of static and dynamic regions (see Fig.4). The images are merged into a complete novel view image by superimposing dynamic regions onto a static region.

Each image is generated by different technique. The morphing process uses an approximate 3-D model of a environment and computing visual hull process estimate ranges between a viewpoint and dynamic objects. Therefore, our method can synthesize the novel view images of each region like positioning 3DCG objects.
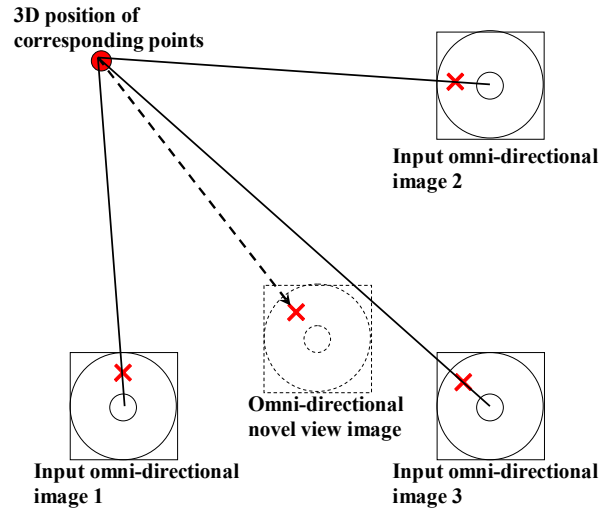

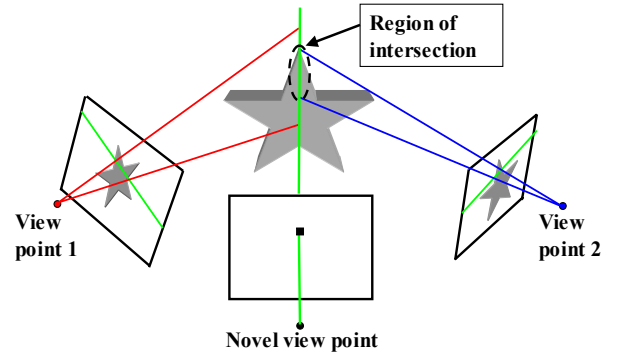Fig.2. Projection of corresponding points to a novel view image.

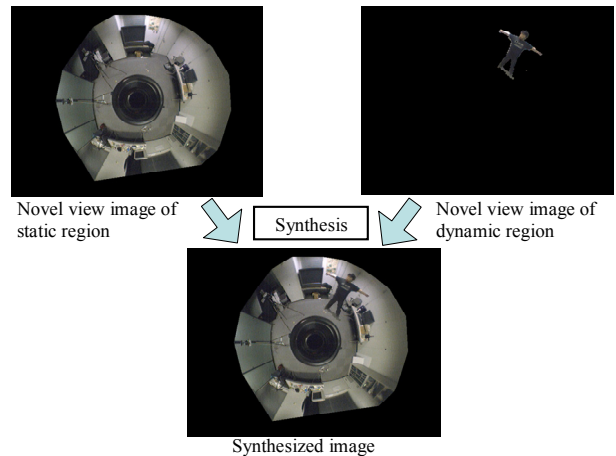Fig.3. Estimation of penetration of visual hull.

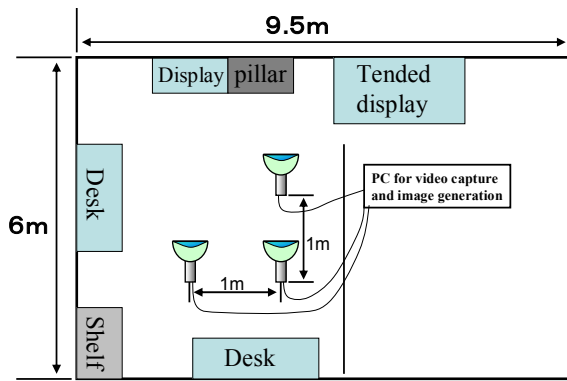Fig.4. Synthesis of static and dynamic regions.

Fig.5. The environment of experiment.

III. EXPERIMENT

We have implemented a prototype system which generates novel view images from live videos using the proposed method. We have carried out experiments with an indoor environment (see Fig.5). In experiments, we have used three omni-directional cameras connected by IEEE1394 to a PC (Pentium4-3.2GHz) for video capture and image generation. The omni-directional cameras can capture the omni-directional video streams (640x480, 15fps).

In the present experiment, the point correspondences among the static scene images have been given manually in advance and have never been renewed. In addition to the reduction of the cost in computing visual hulls described in Section II, we have limited the regions for computing the visual hulls to the view of user.

Fig.6 shows the positions of novel viewpoints, the appearances of experiment, the generated omni-directional novel view images, and the planar perspective images presented to the user. Each planar perspective image can be generated in 200ms. The displacement between the real dynamic objects and the virtual dynamic objects of visual hull is about 50cm. We consider that the displacement is caused by some errors in the positions and postures of cameras and insufficient correspondences given for morphing.
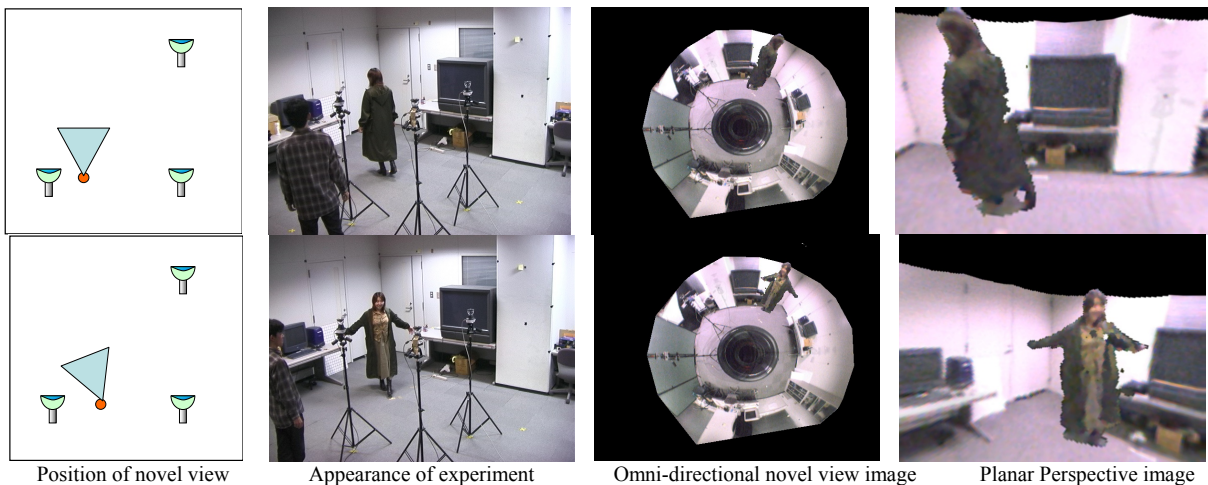
IV. CONCLUSION

We have proposed a method which can generate an omni-directional novel view image of a dynamic scene with a wide view integrating morphing and visual hull techniques. In experiment, the prototype system has been proven to generate novel view images from live videos in real-time. The future work includes the improvement of generated image quality and the realization of more rich presence.

REFERENCES

[1] H.Saito, S.Baba, and T.Kanade, "Appearance-Based Virtual View Generation form Multicamera Videos Captured in the 3-D Room," IEEE Trans. on Multimedia, vol. 5, no.3, pp. 303–316, 2003.

[2] N.Inamoto and H.Saito, "Fly-Through Observation System for 3D Soccer Movie Based on Viewpoint Interpolation," Journal of the Institute of Image Infromation and Television Engineers, vol.58, no.4, pp.529–539, 2004.

[3] T.Koyama, I.Kitahara, and Y.Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," Proc. 2nd IEEE / ACM Int. Symposium on Mixed and Augmented Reality (ISMAR03), pp. 178–187, 2003.

[4] K.Tomite, K.Yamazawa, and N.Yokoya, "Arbitrary Viewpoint Rendering from Multiple Omnidirectional Images for Interactive Walkthroughts," Proc. 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002), vol.3, pp.987-990, 2002.

[5] Y.Onoe, K.Yamazawa, H.Takemura, and N.Yokoya, "Telepresence by real-time view-dependent image generation from omni-directional video stream," Computer Vision and Image Understanding, vol.71, no.2, pp.154-165, 1998.

[6] T.Ishikawa, K.Yamazawa, T.Sato, S.Ikeda, Y.Nakamura, K.Fujikawa, H.Sunahara, and N.Yokoya, "Networked telepresence system using web browsers and omnidirectional video streams," Proc. SPIE Electronic Imaging, vol.5664, pp.380-387, 2005.

[7] S.Morita, K.Yamazawa, and N.Yokoya, "Networked video surveillance using multiple omnidirectional cameras," Proc. 2003 IEEE Int. Conf. Symposium on Computational Intelligence in Robotics and Automation, pp.1245-1250, 2003.

[8] T.Matsuyama and T.Takai, "Generation, Visualization, and Editing of 3D Video," Proc. 1st Int. Symposium on 3D Data Processing, Visualization and Transmission, pp. 234–245, 2002.

[9] W.Matusik, C.Buehler, R.Raskar, and L.McMillan, "Image-Based Visual Hulls", Proc. ACM SIGGRAPH2000, pp.369-374, 2000.

| Position of novel view | Appearance of experiment | Omni-directional novel view image | Planar Perspective image |

Fig.6. Appearance of experimental enviromnent and generated images.