K-010

Development of a 360 Video Player for Head-Mounted Display to Enable Comparison Between Before and After Incident

Hiroyuki Kawakita Toshiyuki Niida and Kensuke Hisatomi

1. Introduction

We have reported devastation of disasters in detail by VR (Virtual Reality) and AR (Augmented Reality) technology [1]. Viewers are able to experience horribleness of the devastation through an HMD (head-mounted display) as if they were there by recording the 360 videos immediately after the occurrence. It is important for them to know the situation not only immediately after it but also during subsequent reconstruction to prepare for the next disaster. We have recorded the 360 videos of such situations both immediately after it and during reconstruction a few months later at the same position. We expect that viewers can understand extent of the damage and the progress of the reconstruction by comparing the difference between the two videos at the same position. In fact, if a viewer watches each video sequentially, it is not easy to compare them because he or she needs to remember where everything was (hereinafter, referred to as “Memory-problem”). Therefore, we developed a 360 player in which a viewer can switch two videos with wipe transition interactively. Using the player, it is easy for viewers to compare between before and after incident.

2. System design

2.1 Implementation approach

To solve the above Memory-problem, we need some kind of mechanism for watching two videos at the same time. In other words, two videos need to be combined into a single video. Depending on the timing of the combination, there are two methods as follows:

- Pre-combination method
  Combine into a single video
  (Pro) Viewers can be easy to notice the difference at a specific position and time of the 360 video that the producer indicate.
  (Pro) A built-in 360 video player on an HMD can be used.
  (Con) Editing is required for each content
  (Con) Memory-problem still remains in most areas but the specific position and time.
- Real-time-combination method
  Combine on the player in a viewer’s playing
  (Pro) Memory-problem does not occur at any position and time. (It is suitable for 360 videos)
  (Pro) It keeps viewers’ interest in 360 videos. (Because of increasing viewer appreciation of 360 video by interaction [2])
  (Con) A dedicated application with an HMD needs to play.
  (Con) High-performance computing resources are required to play.

Japan Broadcasting Corporation (NHK)

Since the latter is more effective and worth a technological challenge, we decided to develop a player application.

2.2 Combination method

In the real-time-combined method, how to combine two playing 360 videos can be regarded as how to switch them by viewer operation. With reference to the transitions used as video switching in the video editing, we can use some switching methods as follows:

- Cut-in
  Switching the whole image of them instantly by pushing a button.
  (Pro) Easy to understand and use.
  (Con) The wide-field visual change is too sudden and causes a lack of cognitive continuity.
- Alpha Blending
  Switching them gradually by adjusting opacity of the front video.
  (Pro) Easy to compare.
  (Con) Neither video can be seen clearly whenever the front video is translucent.
- Wipe
  Switching them with wipe-transition by controlling position of the border.
  (Pro) Easy to focus on the difference.
  (Con) The local misalignment at the border is emphasized.
- Paint
  Switching them as if to paint it over.
  (Pro) Playable like a treasure hunt.
  (Con) Too imprudent for some kinds of content.

There are advantages and disadvantages with the methods, it depends on content to choose which one to use.

As an example of the Alpha Blending method, Okada et al.[3] proposed an AR application on smartphone to compare between the real scene in front of a user and the past photo of the same scene. Viewing the scene through the smartphone’s rear camera from a specific standing position, the user can notice the difference interactively by changing the transparency of the overlaid past photo.

As mentioned in chapter 1, since this time we assume disaster content, every disadvantage above is undesirable. Nonetheless, there is room for improvement in the Wipe method if the border is obscured to make the misalignment smooth. Therefore, as shown in Fig.1, we adopted the Wipe method in which the two 360 videos (two hemispheres) are combined with a vertical plane as a border, which viewers can move horizontally by his/her interaction with the controller. The border is obscured by the decoration like a white light beam. The viewer can compare situation between immediately after the disaster and during

Fig.1 360 video switching with Wipe method

reconstruction a few months later at the same position without the Memory-problem.

3. Monocular application

In this chapter, we describe development of the application which displays the image from the same video toward each eye without binocular parallax in an HMD (hereinafter, referred to as “monocular application”). On the other hand, the next chapter shows the application which displays the image from the different video toward each eye with binocular parallax (hereinafter, referred to as “Binocular application”). In comparison to the binocular application, the features of the monocular application are that simple 360 camera is available for shooting and that less computer resources are required for rendering.

3.1 Implementation

In this section, we describe concrete implementation of the application according to the design described in the previous chapter. As shown in Fig.2, using a game engine Unity, the two videos were pasted on the inner surfaces of the two spheres respectively in virtual space. The spheres were placed so as to sandwich a viewer (= a viewpoint of an HMD) between centers of them in the horizontal plane. Since each center revolves around the viewer in a circular motion in the horizontal plane by the controller, the hemisphere-videos appear to be switched by the Wipe method. Note that each sphere does not rotate.

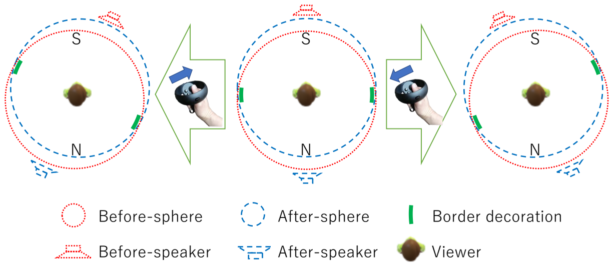

Fig.2 Implementation of Wipe method

Each audio was designed to be output from the virtual speaker which was placed at the top of each apparent hemisphere dome. Each speaker also revolves with the sphere. Each sound was recorded omnidirectional, therefore no matter how we structure the sound field, it is not natural.

Table 1 shows the development environment and the content parameter.

3.2 Results

3.2.1 Load on computer

Table 2 shows the rendering environment and the load on the computer. The load of the process was recorded for 30 seconds in a steady state and the maximum and minimum values were shown. It was observed that the video files had been read out from the hard drive and that the two video files were all extracted in the memory (RAM). Although decoding two 4K videos at the same time takes a load on the CPU, any dropped frame or other problems was not found in an HMD.

3.2.2 Brief evaluation

As a brief evaluation, we asked about 20 people to experience the application and give us their impressions. The content was a series of short clips of 30 seconds each at three locations in the Western Japan’s rain disaster[4] on July, 2018. The viewer can compare situations between immediately after the occurrence and during reconstruction a few months later at the same position.

The main impressions were as follows:

- Wiping shows horribleness of the damage.

Table.1 Development environment and content parameter

360 camera		Insta360 Pro
Video editing tool		Premiere PRO
		Adobe After Effects (Mocha AE CC)
Application authoring tool		Unity 2019.2.21f1
Video 1	Resolution	4K x 2K (px)
	Frame rate	30 (fps)
	Video codec	H.264 (mp4)
	Bit rate	39.5 (Mbps)
Video 2	Resolution	4K x 2K (px)
	Frame rate	30 (fps)
	Video codec	H.264 (mp4)
	Bit rate	63.1 (Mbps)

Table.2 Rendering environment

Target HMD		Oculus Rift S
Rendering computer	CPU	AMD Ryzen 9 3950X 16-Core
	GPU	NVIDIA GeForce GTX 1650 SUPER
	Hard drive	Crucial CT500P1SSD8
	RAM	16.0 (GB)
	OS	Windows 10 Home 64bit
Load on the computer	CPU usage	17-32 (%)
	GPU usage	38-43 (%)
	Hard drive usage	0.1 (MB/s)
	Memory usage	1,790-1,850 (MB)

- I see the houses that were destroyed by the flood have been torn down and mostly vacant lots.
- The road is like a river!
- I like the sound that goes around.
- More resolution of an HMD would be nice.

With the interaction of the Wipe method, even the person who shot and edited the footage could notice minor differences that had not been noticed before. It is assumed that the minor differences can be noticed due to Memory problem being solved.

4. Binocular application

4.1 Implementation

The binocular application requires additional video streams to display different images for the left and right eyes. To set the same resolution as the monocular application, four 4K x 2K video inputs are required. However, due to the high CPU load caused by the decoding process, we found that the environment shown in Table 2 could not playback properly owing to frame dropping. Therefore, by re-encoding the H.264 video to HAP [5], the decoding load on the CPU was transferred to the GPU and drive access. HAP video can be decoded in GPU, and their total file size is much larger than H.264. Table 3 shows the content parameter.

4.2 Results

4.2.1 Load on computer

Table 4 shows the load on the computer. The target HMD and the rendering environment is the same as the one in the monocular application. Compared to the result in Table 2, it was observed that load on CPU was transferred to GPU and drive access. Since the total file size of the videos was much larger than the RAM capacity, the amount of hard drive access was high as much as streaming. Any dropped frame or other problems was not found in an HMD.

4.2.2 Brief evaluation

As in subsection 3.2.2, we asked about 20 people to experience the application and give us their impressions. The content was also similar to that described in 3.2.2. Since most of the people did not experience the monocular application, most impressions were similar to those in 3.2.2. The main impressions by those who had experienced both applications were as follows:

- I like the binocular version because it feels real.
- This is more natural.
- It makes nearby objects feel closer.
- I wondered why I felt my vision was narrower than monocular application.

Table.3 Content parameter

Video 1	Resolution	4K x 4K (px)
	Frame rate	30 (fps)
	Video codec	Hap (mov)
	Bit rate	1,209 (Mbps)
Video 2	Resolution	4K x 4K (px)
	Frame rate	30 (fps)
	Video codec	Hap (mov)
	Bit rate	1,353 (Mbps)

Table.4 Load on the computer (Binocular)

CPU usage	8-10 (%)
GPU usage	48-55 (%)
Hard drive usage	316-333 (MB/s)
Memory usage	342-344 (MB)

5. Discussion

5.1 Comparison between monocular and binocular

The binocular application gave better impressions than the monocular application, and as far as we have heard this time, there were no negative comments about it. Although there were concerns that the inconsistency of stereopsis at the border might lead to discomfort, it had virtually no effect. The binocular application is useful when the producer wants to convey a sense of realism, such as when there is an object a few meters away from the viewpoint. In this case, the effectiveness of the binocular application was demonstrated in the disaster content. A 360 camera with stereo vision and some rendering computer resources are needed for the binocular application. In order to make effective use of the limited resources, it is possible to devise ways to spread the load, as discussed in section 4.1.

5.2 Sound field structure

In this case, the sound was recorded omnidirectional. Rotating virtual speakers successfully gave the impression that the world was spinning. However, it is an unnatural method because the sound source cannot be localized. It might have been better if it had been recorded in ambisonics. In that case, it would be preferable to divide it in half and to use it with the corresponding hemisphere. Real-time sound rendering is necessary to change the division method to match the wipe.

6. Conclusion

The essence of VR is to transcend time and space. The applications we developed takes viewers to the scene of a disaster, and they wipe to switch the time. By learning more about the devastation of the disaster and how they are recovering, we aim to raise their awareness of disaster prevention. With the current application, they can only switch between two times at certain points, but we would like to develop a more flexible way to transcend time and space.

References

[1] NHK VR/AR, https://www.nhk.or.jp/vr/
[2] Y. Sakakibara, R. Tanaka, T. Narumi, T. Tanikawa, M.Hirose, “Increasing User Appreciation of Spherical Videos by Finger Touch Interaction”, Proc. of 18th Human–Computer Interaction International (HCII 2016), Part I, pp.546-555 (2016)
[3] N. Okada, J. Imura, T. Narumi, T. Tanikawa, and M. Hirose, “Manseibashi reminiscent window: On-site ar exhibition system using mobile devices”, International Conference on Distributed, Ambient, and Pervasive Interactions, LNCS vol. 9189, pp. 349-361. (2015)
[4] Western Japan’s rain disaster, https://www3.nhk.or.jp/nhkworld/en/news/backstories/178/
[5] HAP, https://hap.video/