SMOOTHLY SWITCHING METHOD OF ASYNCHRONOUS MULTI-VIEW VIDEOS USING FRAME INTERPOLATION

Aoi Harazaki   *Hidehiko Shishido*   *Yoshinari Kameda*   *Itaru Kitahara*

University of Tsukuba   University of Tsukuba   University of Tsukuba   University of Tsukuba
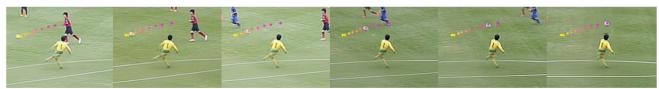
Figure 1: Perform asynchronous multi-viewpoint photographing by using an inexpensively installable capturing system in a large-scale space. In the bullet-time image using the captured image, the observation position on the image of the subject is blurred at the viewpoint switching due to the influence of the synchronization shift. Therefore, by adapting the proposed method, we generate a bullet time image in which the blur of the observation position of the object is reduced.

ABSTRACT

This paper proposes a method to generate smoothly view-point switching by reducing the flickering artifact observed at Bullet-Time generated from asynchronous multi-view videos using frame interpolation processing. When we asynchronously capture multi-view videos of an object moving at high velocity, deviations occur in the observed position on Bullet-Time. We apply frame interpolation technique to reduce the problem. By selecting a suitable interpolated images that produce the smallest movement of the subject's observed position, smoothly view-point switched Bullet-Time video is generated.

Index Terms —3D free-viewpoint video, bullet-time, frame interpolation, morphing, Asynchronous video

1. INTRODUCTION

3D free viewpoint video technology that is capable of observing the subject from various angles has been actively researched [1]-[4]. In the video, it is possible to observe the subject from the arbitrary position wherever the observer wants to see. It can correctly express spatially relationship between captured subjects rather than ordinary video observing using conventional monocular cameras. So, image industry such as movies and the application fields such as sports science are paying attention on it.

In the conventional 3D free viewpoint video technology, there are problems such as degradation of the quality of the generated video due to the estimation error of the 3D shape and high calculation cost. The influence of the estimation error can be reduced by approximating the object shape to a 2D billboard plane [5], however the image quality degradation due to the subject tracking or image segmentation error still remains.

One of the most important purposes of 3D free viewpoint video is generating motion parallax (i.e., visual effect that the viewpoint is virtually moving). By focusing on this point, we research about the bullet-time video that reproduces such visual effect by continuously switching the images captured by multiple cameras surrounding the subject. [6, 7, 8]. By presenting the captured image almost as it is, the bullet-time can present a high-quality video with little image degradation due to processing error.

However, there are some issues to install the bullet-time to the actual sports field, such a soccer stadium. For example, when we capture multi-view images for 3D computer vision processing or Free Viewpoint Television (FTV) [9], normally, cables for signal distribution are connected to all cameras and synchronous signal is distributed from the synchronizing signal distributor. When the number of the capturing camera or the size of the target space is larger, the cost for setting becomes higher. In response to such problems, synchronous imaging methods that can be applied to large-scale spaces have been proposed by generating external synchronization signals from GPS signals and distributing them to multi-view cameras [10]. However, such special device makes expensive the installing cost. In order to realize reasonable image-capturing at practical shooting environments, a method synchronizing multi-view asynchronous images using audio information has been proposed [11]. In our target environment such as a soccer stadium, however, there is serious sounds echo effect (reflection sound), that makes difficult the synchronization process. Moreover, if the system can archive frame-level synchronization, asynchronous of shutter-level still remains, since the internal clocks of the cameras are not synchronized.

As the mentioned above, if we consider about accurate 3D computer vision processing, completely synchronized multi-view videos taken with huge amount of labor and special equipment are required. On the other hand, the bullet-time video can be generated with asynchronous videos, since it does not need accurate 3D information. However, when we capture fast moving objects, flickering artifact (deviations of the observed position) is observed in bullet-time video.

In this research, we aim to reduce the deviation by using frame interpolation technique for generating bullet-time video smoothly switching the viewpoint. In each multi-view video, frame interpolation processing is applied with respect to the frame before and after the frame of interest. Generally, frame interpolation processing is realized by image morphing that generates a series of image groups whose appearance changes continuously from one shape to the other using a process of synthesizing the appearance of the middle between two images [12]. In the typical morphing method [13], correspondence information between two input images is manually given, however when video data composed of 30 or more frames per second is to be processed, manually inputting required correspondence is not feasible. In order to solve this problem, we employ Regenerative

Morphing [14] that can automatically generate interpolated images even if there is no correspondence information between input images. However, when the target object moves fast, the position of the target object between consecutive front and rear frames is greatly different. As the result, it is difficult to generate an interpolation image in which the object is clearly observed at the desired position. In this research, we deal with this problem by generating an initial interpolation image that includes the cue of visual information to generate an appropriate interpolation image from a simple correspondence.

2. SMOOTH MULTI-VIEW IMAGE SWITCHING PROCESSING

The proposed method is mainly composed of two processes. The first is processing for generating a plurality of interpolated images between consecutive before and after frames using frame interpolation processing at each viewpoint. The second is to choose an appropriate image with a small change in observation position of the subject out of the generated interpolation images, so that the sense of discomfort at the time of viewpoint switching in the bullet-time is reduced.

2.1 Frame interpolation using Regenerative Morphing

For frame interpolation processing, Regenerative Morphing is applied. Figure 2 shows an interpolated image T_n between the input images S_1 and S_2. Figure 2 (a) shows an image generation by blending S_1 and S_2. In the target interpolation image T_n, interpolation image generation is realized as an energy optimization problem defined by a bidirectional similarity function [15] using consistency between adjacent and previous interpolation images and similarity of input images. Figure 2 (b) is an interpolated image generated by initial image. Here, the blended image ((a)middle) is used for the initial image for the interpolation.

Since, in the initial image for the interpolation, there is no visual cue information of the target object around the center of the image where the target object should be observed, in the interpolation result ((b) middle), the appearance of the target object is not well generated. This is because, the simple blending result of the input image is the initial interpolation image, when the subject moves at high speed, information of the target object's appearance does not exist at an appropriate position in the initial interpolation image.
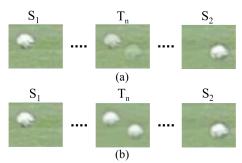

Figure 2: Interpolated image of Regenerative Morphing

We solve this problem by giving only one set of correspondence information of subjects between input images. As shown in Figure (a) middle, by generating motion blur from given correspondence information, the cue of appearance is given at the position where the target object should be observed in the initial interpolation image. The area of motion blur is added depends on the size of the target object. As shown in Figure 3, when the size of the object is H, the distance is D, the focal length is d, the size of the image sensor is x, and the number of pixels is X, the size h of the object on the image is given by Equation(2).
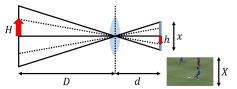

Figure 3: Size of subject on image

$$H : h\frac{x}{X} = D : d \qquad (1)$$

$$h = Hd \times \frac{X}{Dx} \qquad (2)$$

In order to generate a plurality of interpolation images with different interpolation distances between input images, the blending ratio of the subject and the background is determined according to the interpolation distance between the two input images by using the normal distribution shown in Equation(3). Let x be the position between the input images, let μ be the ratio of the interpolation distance to the input image and σ^2 the variance value.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \qquad (3)$$

Figure 4 (a) shows initial interpolation image applying the motion blur effect. Figure (b) is an interpolation images generated using the initial interpolation images. In the interpolation image T_n, it is found that the target object is clearly observed at the desired position.
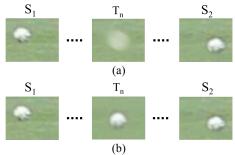

Figure 4: Interpolated image of Regenerative Morphing

2.2 Selection method of generated interpolation image

We describe a method to choose an appropriate interpolation image that makes the change of observation position of the target object smooth when the viewpoint is moving. Figure 7 shows the movement of the target objects during the viewpoint movement. When switching the multi-view camera from left to right, it can be confirmed that the position where the target object (player) is observed changes from right to left (Figure 7 left). Since the movements of the blue and red players are sufficiently slow comparing with the camera shutter speed, the influence of the synchronization deviation is small, and as a result, the observed position moves along with a certain curve. On the other hand, the soccer ball moves much faster so that it is greatly influenced to the synchronization deviate, as the result, the observed position deviates. In Figure 7 center, it can be confirmed that the ball has not moved in a certain direction on the curve.

We assume that smooth viewpoint switching is realized by suppressing the deviation of observation position.

In this section, we describe a method of suppressing the change of the observation position of the subject and selecting an image that will be smoothly switched when we switch images. First of all, in all of the input image and the interpolation image of the multi-view camera, the observation position of the fast-moving target object on the image is acquired. Next, as shown in figure 5, the angle θ_n between the following two vectors is calculated. A vector $\boldsymbol{a}_n$ obtained from the coordinate $P_n(x_n, y_n)$ on the image of the camera n and the coordinate $P_{n+1}(x_{n+1}, y_{n+1})$ on the image of the camera n+1. A vector $\boldsymbol{a}_{n+1}$ obtained from the coordinate $P_{n+1}(x_{n+1}, y_{n+1})$ on the image of the camera n+1 and the coordinate $P_{n+2}(x_{n+2}, y_{n+2})$ on the image of the camera n+2. The accumulation of this angle is calculated by all the cameras（Cam1,…,CamN）using Equation(4). Similar processing is performed for all combinations of the input image and the interpolation image of the multi-view cameras, and an image that has the minimum accumulation value is chosen. The smaller the angle between the two vectors is, the more the coordinates of the three points exist in a constant curve shape. The smaller the Equation(4) is, the smaller the change in the observation position of the object is, and smooth image switching is performed.

$$\sum_{n=1}^{N-1} \theta_n = \sum_{n=1}^{N-1} \cos^{-1}\left(\frac{\boldsymbol{a}_n \cdot \boldsymbol{a}_{n+1}}{|\boldsymbol{a}_n||\boldsymbol{a}_{n+1}|}\right) \qquad (4)$$

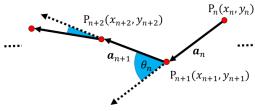$$\boldsymbol{a}_n = (x_{n+1} - x_n, y_{n+1} - y_n) \qquad (5)$$

Figure 5: Changes in observation coordinates

3. EXPERIMENT AND RESULTS

We conducted on video capturing experiment in the soccer game of the junior youth held at the Japanese football stadium (Kashima soccer stadium). As shown in Figure 6 and 7, ten video cameras (Canon EOS 5D Mark II, 1,920 pixels × 1,080 pixels at 30 frames per second) are installed in the audience seat behind the goal to capture the near penalty area. The convergence angle of neighboring cameras is about 10 degrees. Image processing is performed using a computer equipped with CPU: Intel Core i7 3.4GHz, GPU: Intel HD Graphics Family, memory: 8.0G.

Figures 8 and 9 show the results of our proposed method. At the center of Figure 8 is the result of comparative bright combining processing applied to all camera sequence image. Likewise, on the right of Figure 8, it is the result of comparative bright combining processing applied to the interpolated image generat-

ed by our proposed method. In the middle of Figure 8, the position of the soccer ball moves forward and backward, whereas in the proposed method on the right of Figure 8 it can be confirmed that it moves in a constant direction along a smooth curve. As shown in Figure 1, it can be confirmed that when the result of application of the proposed method is seen frame by frame, it is similarly moving in a certain direction along a smooth curve.

Furthermore, we confirmed the effectiveness of the proposed method for another sequence image. Figure 9 (a) shows a bullet time image by switching multi-view images taken, and Figure 9 (b) shows a bullet time image using our proposed method. It shows the observation position of the soccer ball in a point changing from yellow to red. When switching the image, the position of the soccer ball moves forward and backward in Figure 9 (a), whereas in Figure 9 (b) it moves in a constant direction along a smooth curve. Especially in Figure 9 (a), the trajectory of the observation position is greatly deviated, on the other hand, in Figure 9 (b) it is found that it has been eliminated. By minimizing the sum of the angles obtained from the three observation coordinates, all observed coordinates are corrected. And it turns out that smoother switching is realized.

In this way, we confirmed the effectiveness of the proposed method as a result of experiments on various fast-moving sequences.
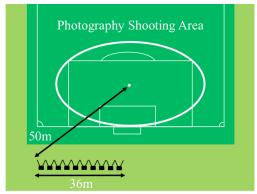

Figure 6: Shooting environment outline

4. SUMMARY

We proposed a method to generate a smooth viewpoint switching image by reducing the deviation observed in the bullet-time image generated from the asynchronous multi-view video using frame interpolation processing. In each multi-view camera between consecutive front and rear frames, the interpolation processing result was improved by using the initial interpolation image weighted based on the normal distribution when frame interpolation processing is performed on the subject moving at high velocity. We also made it possible to generate a bullet time image with smooth viewpoint switching by selecting an interpolation image that minimizes the change of the observation position of the subject when moving the viewpoint.

Figure 7: All captured camera sequence images

(a) Player position   (b) Previous method (ball)   (c) Proposed method (ball)

Figure 8: (a) This is the result of comparative dark combining processing applied to all camera sequence image. When switching the multi-view camera from left to right, it can be confirmed that the position where the subject (player) is observed changes from right to left. (b) This is the result of comparative bright combining processing applied to all camera sequence image. (c) This is the result of comparative bright combining processing applied to the interpolated image generated by the proposed method.
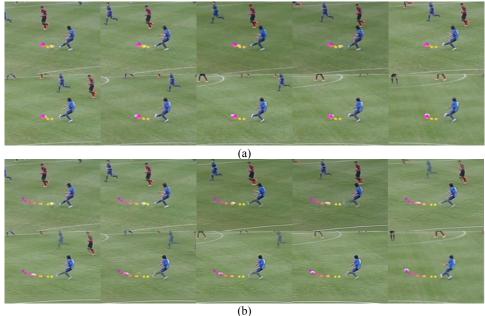

(a)

(b)

Figure 9: (a) shows a bullet time image by switching multi-view images taken, and (b) shows a bullet time image using the proposed method.

5. REFERENCES

[1] J. Carranza, C. Theobalt, M. A. Magnor and H.-P. Seidel, "Free-Viewpoint Video of Human Actors," ACM Transactions on Graphics, Vol.22, No.3, pp.569-577, 2003.

[2] T. Kanade, P. Rander and P. J. Narayanan, "VirtualizedReality: Constructing Virtual Worlds from Real Scenes," IEEE MultiMedia, Vol.4, No.1, pp.34-47, 1997.

[3] Y. Liu, Q. Dai and W. Xu, "A Point-Cloud-Based Multiview Stereo Algorithm for Free-Viewpoint Video," IEEE Transactions on Visualization and Computer Graphics, Vol.16, No.3, pp.407-418, 2010.

[4] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe,A. Kirk, S. Sullivan, "High-Quality Streamable Free-Viewpoint Video," ACM Transactions on Graphics, 34(4), 2015.

[5] T. Koyama, I. Kitahata and Y. Ohta, "Live Mixed- Reality 3D Video in Soccer Stadium," Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp.178-186, 2003.

[6] N.Akechi, I.Kitahara, R.Sakamoto and Y.Ohta, "Multi-resolution bullet-time effect," SIGGRAPH-ASIA,2014.

[7] T. Kanede et al.,"EyeVision," http://www.ri.cmu.edu/events/sb35/tksuperbowl.html, 2001.

[8] K. Tomiyama and Y. Iwadate, "Development of Multi-View HDTV Image Generation System," The Journal of the Institute of Image Information and Television Engineers, Vol.64, No.4, pp.622- 628, 2010.

[9] M. Tanimoto "Free-Viewpoint TV," IEEE Signal Processing Magazine, Vol.28, No.1, pp.67-76, 2011.

[10] I. Kitahara, H. Saito, Y. Ohta,T .ono, T. Kanade, "Large-scale Virtualized Reality," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Technical Sketches, 2001.

[11] RedGiant,"Plural Eyes," https://www.redgiant.com/products/pluraleyes/.

[12] T. Beier and S. Neely, "Feature-Based Image Metamorphosis," ACM SIGGRAPH Computer Graphics, Vol.26, No.2, pp.35-42, 1992.

[13] S.Y.Lee and K.Y.Chwa, "Image Morphing Using DeformationTechniques," The Journal of Viuallizaiion and Computer Animation, Vol.7, pp.3-23, 1996.

[14] Shechtman E, Rav-Acha A, Irani M, Seitz S, "Regenerative Morphing," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San-Francisco CA, June 2010.

[15] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. "Summarizing visual data using bidirectional similarity," In CVPR,Anchorage, AK, USA, 2008.