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

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

University of Tsukuba

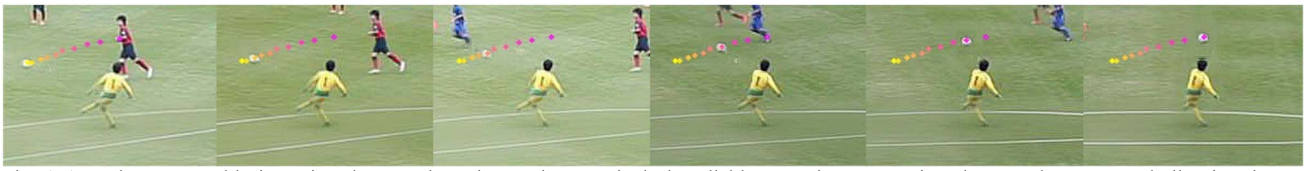


Fig. 1 Asynchronous multi-viewpoint photography using an inexpensively installable capturing system in a large-scale space. In bullet-time image using the captured image, observation position on subject's image is blurred at viewpoint switching due to influence of synchronization shift. Therefore, by adapting the proposed method, we generate a bullet-time image that reduces blur of object's observation position.

ABSTRACT

This paper proposes a method that generates viewpoint smooth switching by reducing the flickering artifact observed at bullet-times 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 at the bullet-times. We apply a frame interpolation technique to reduce the problem. By selecting suitable interpolated images that produce the smallest movement of the subject's observed position, we smoothly generate viewpoint switched bullet-time video.

Index Terms: 3D-free-viewpoint video, bullet-time, frame interpolation, morphing, asynchronous video

1. INTRODUCTION

3D-free-viewpoint video technology, which can observe a subject from various angles, is actively being researched [1]-[4]. In video, the subject can be observed from arbitrary positions that reflect what the observer wants to see. Such advanced video technology can correctly express spatial relationships between captured subjects rather than ordinary video that observes with conventional monocular cameras. Such industries as movies and application fields, including sports science, are paying attention to it.

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

One of the most critical purposes of 3D-free-viewpoint video is generating motion parallax, which is a visual effect where the viewpoint is virtually moving. By focusing on this point, we research bullet-time video that reproduces such visual effects by continuously switching the images captured by multiple cameras surrounding the subject [6, 7, 8]. By presenting the captured image almost as it is, bullet-time presents high-quality video with little image degradation from processing error.

However, some issues remain concerning the installation of bullet-time into actual sports facilities, such as soccer stadiums. For example, when we capture multi-view images for 3D computer vision processing or Free Viewpoint Television (FTV) [9], cables for signal distribution are normally connected to all of the cameras, and a synchronous signal is distributed from a synchronizing signal distributor. When the number of capturing cameras or the size of the target space is increased, the establishment cost is also raised. In response to such problems, synchronous imaging methods, which 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 devices increase the installation cost. For reasonable image-capturing in practical shooting environments, a method has been proposed that synchronizes multi-view asynchronous images using audio information [11]. In our target environment, such as a soccer stadium, however, serious echo effects (reflection sound) complicate the synchronization process. If a system can achieve frame-level synchronization, shutter-level asynchronous remains, since the internal clocks of the cameras are not synchronized.

As mentioned above, if we consider accurate 3D computer vision processing, completely synchronized multi-view videos are required that are taken with a huge amount of labor and special equipment. On the other hand, 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 reduce the deviation using a frame interpolation technique for generating bullet-time video that smoothly switches the viewpoint. In each multi-view video, frame interpolation processing is applied with respect to the frame before and after the frame of interest. Frame interpolation processing is generally achieved by image morphing that generates a series of image groups whose appearance changes continuously from one shape to another using a process that synthesizes the appearance of the middle between two images [12]. In the typical morphing method [13], the correspondence information between two input images is manually given; however, when video data composed of 30 or more frames per second are processed, manually inputting that requires correspondence is not feasible. To solve this problem, we employ regenerative morphing [14], which can automatically generate interpolated images even without correspondence information between input images. However, when the target object moves quickly, its position between consecutive front and rear frames is greatly different. As a result, it is difficult to generate interpolation images in which the object is clearly observed at the desired position. In this research, we address this problem by generating an initial interpolation image that includes the visual information cue to generate an appropriate interpolation image from a simple correspondence.

2. SMOOTH MULTI-VIEW IMAGE SWITCHING PROCESSING

Our proposed method is mainly composed of two processes. The first consecutively generates a plurality of interpolated images between before and after frames using frame interpolation processing at each viewpoint. The second chooses an appropriate image with a slight change in the subject's observation position out of the generated interpolation images to reduce the sense of discomfort at the time of the viewpoint switching in the bullet-time.

2.1 Frame interpolation using regenerative morphing

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

In the initial image for the interpolation, since there is no visual cue information of the target object around the image's center where the target object should be observed, in the interpolation result (Fig. 2(b), middle), the target object's appearance 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, and information of the target object's appearance does not exist at an appropriate position in the initial interpolation image.

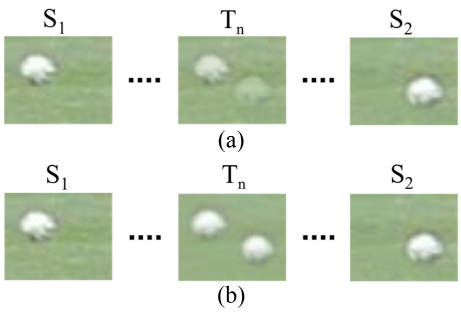


Fig. 2 Interpolated image of regenerative morphing

We solve this problem by giving just one set of correspondence information of the subjects between the input images. By generating motion blur from the given correspondence information, an appearance cue is given at the position where the target object should be observed in the initial interpolation image (Fig. 2(a), middle). The motion blur area that is added depends on the target object's size. As shown in Fig. 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, the number of pixels is X, and size h of the object on the image is given by Eq. (2):

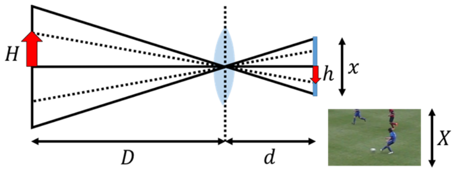


Fig. 3 Size of subject on image

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

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

To generate a plurality of interpolation images with different interpolation distances between input images, we determined the blending ratio of the subject and the background based on the interpolation distance between the two input images using the normal distribution shown in Eq. (3). Let x be the position between the input images, let μ be the ratio of the interpolation distance to the input image, and let σ^2 be 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 the initial interpolation image after applying the motion blur effect. Fig. 4(b) is an interpolation image generated using the initial interpolation images. In interpolation image T_n, the target object is clearly observed at the desired position.

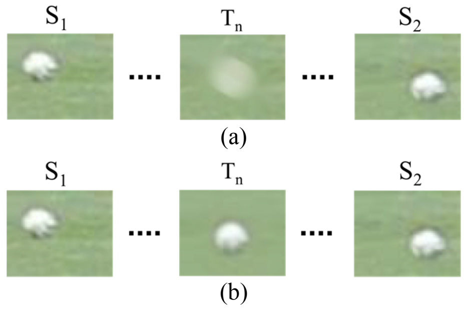


Fig. 4 Interpolated image of regenerative morphing

2.2 Selection method of generated interpolation image

Next we describe a method for choosing an appropriate interpolation image that smoothly changes the observation position of the target object when the viewpoint is moving. Fig. 7 shows the movement of the target objects during the viewpoint movements. When switching the multi-view camera from left to right, we confirmed that the position where the target object (player) is observed changes from right to left (Fig, 7, left). Since the movements of the blue and red players are sufficiently slower than the camera shutter speed, the influence of the synchronization deviation is small, and as a result, the observed position moves along a certain curve. On the other hand, since the soccer ball moves much faster, it is greatly influenced by the synchronization's deviation, and the observed position is deviated. In Fig, 7, center, we confirmed that the ball did not move in a certain direction on the curve. We assume that smooth viewpoint switching is achieved by suppressing the deviation of the 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, in all of the input and interpolation images of the multi-view camera, we acquired the observation position of the fast-moving target object on the image. Next, we calculated angle θ_n between the following two vectors (Fig. 5). Vector $\boldsymbol{a}_n$ was obtained from coordinate $P_n(x_n, y_n)$ on the image of camera n and coordinate $P_{n+1}(x_{n+1}, y_{n+1})$ on the image of camera n+1. Vector $\boldsymbol{a}_{n+1}$ was obtained from coordinate $P_{n+1}(x_{n+1}, y_{n+1})$ on the image of camera n+1 and coordinate $P_{n+2}(x_{n+2}, y_{n+2})$ on the image of camera n+2. The accumulation of this angle is calculated by all the cameras (Cam1,…,CamN) using Eq. (4). Similar processing was performed for all of the combinations of input and interpolation images of the multi-view cameras, and an image is chosen with minimum accumulation values. The smaller the angle is between the two vectors, the more coordinates of the three points exist in a constant curve shape. The smaller Eq. (4) is, the smaller is the change in the observation position of the object, and smooth image switching is performed:

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

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

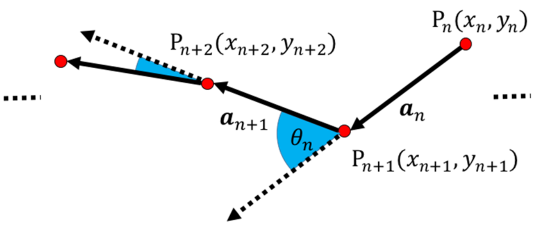


Fig. 5 Changes in observation coordinates

3. EXPERIMENT AND RESULTS

We conducted a video-capturing experiment of a youth soccer match in a large soccer stadium. As shown in Figs. 6 and 7, we installed ten video cameras (Canon EOS 5D Mark II, 1,920 pixels × 1,080 pixels at 30 frames per second) in the seats behind the goal to capture the near-penalty area. The convergence angle of the neighboring cameras was about 10 degrees. Image processing was performed using a computer with the following characteristics: CPU: Intel Core i7 3.4 GHz, GPU: Intel HD Graphics Family, memory: 8.0 G.

Figures 8 and 9 show the results of our proposed method. At the center of Fig. 8 is the result of the comparative bright combining processing that was applied to all of the camera sequence images. On the right of Fig. 8 is the result of the comparative bright combining processing that was applied to the interpolated images generated by our proposed method. In the middle of Fig. 8, the soccer ball position moves forward and backward; in the proposed method on the right of Fig. 8, it moves in a constant direction along a smooth curve. As shown in Fig. 1, we confirmed that when the application result of the proposed method is seen frame by frame, it moves similarly in a certain direction along a smooth curve.

We also confirmed the effectiveness of the proposed method for another sequence image. Fig. 9(a) shows a bullet-time image by switching the multi-view images, and Fig. 9(b) shows a bullet-time image using our proposed method. The observation position of the soccer ball in a point changes from yellow to red. When switching the image, the soccer ball's position moves forward and backward (Fig. 9(a)), but in Fig. 9(b) it moves in a constant direction along a smooth curve. Especially in Fig. 9(a), the trajectory of the observation position is greatly deviated, but that problem has been eliminated in Fig. 9(b). By minimizing the sum of the angles obtained from the three observation coordinates, all of the observed coordinates are corrected. This step realizes smoother switching.

In this way, we experimentally confirmed the effectiveness of our proposed method on various fast-moving sequences.

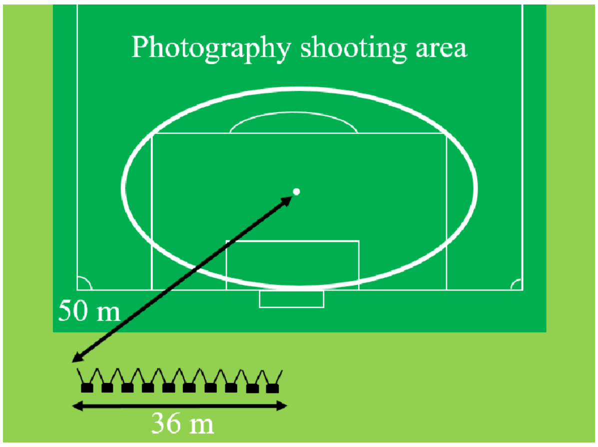


Fig. 6 Shooting environment outline

4. SUMMARY

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

This work was supported by JSPS KAKENHI Grant Numbers 17H01772 and CREST, JST.

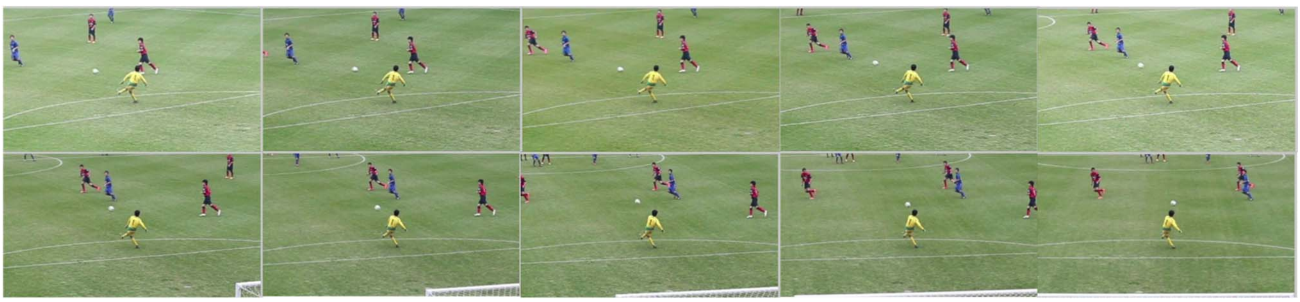
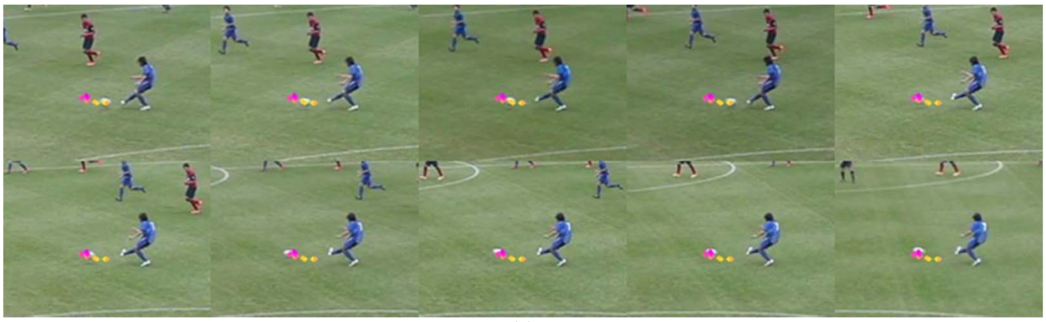


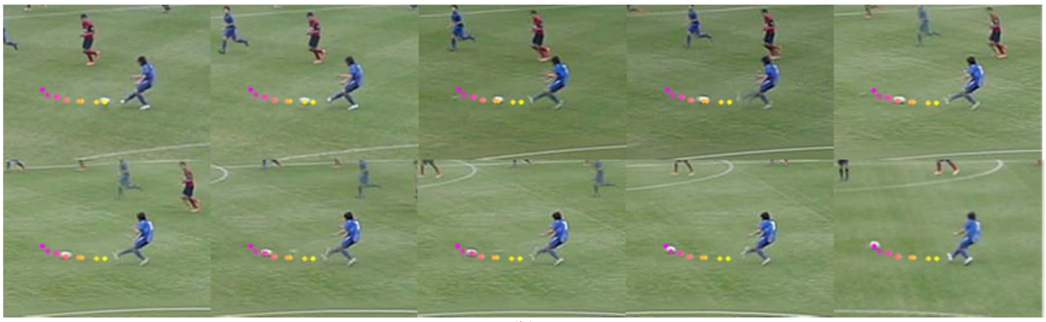
Fig. 7 All captured camera sequence images

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

Fig. 8 (a) Comparative dark combining processing result applied to all camera sequence images. When switching multi-view camera from left to right, we confirmed the position, where the subject (player) is observed, changes from right to left. (b) Comparative bright combining processing result applied to all camera sequence images. (c) Comparative bright combining processing result applied to interpolated image generated by proposed method.



(a)



(b)

Fig. 9 (a) Bullet-time image taken by switching multi-view images and (b) bullet-time image using 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, and 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, and 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 Visualization and Computer Animation, Vol. 7, pp. 3-23, 1996.

[14] Shechtman E, Rav-Acha A, Irani M, and 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.