J. Vis. Commun. Image R. 62 (2019) 68–76

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Smooth switching method for asynchronous multiple viewpoint videos using frame interpolation ☆

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

University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

ARTICLE INFO

Article history:
Received 5 July 2018
Revised 11 March 2019
Accepted 20 April 2019
Available online 3 May 2019

Keywords:
Free-viewpoint video
Bullet-time
Frame interpolation
Morphing
Asynchronous multi-view videos

ABSTRACT

This research proposes a method that generates viewpoint smooth switching by reducing the flickering artefacts that are 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 a high velocity, deviations occur in the observed position at the bullet-times. We apply a frame interpolation technique to smooth this problem. By selecting suitable interpolated images that produce the smallest movement of the subject's observed position, we smoothly generate a viewpoint-switched bullet-time video.

In this paper, we examine the subjective evaluation of the video generated by the proposed method. And we also examine objective evaluation. Therefore, the effectiveness of the proposed method is shown. Furthermore, reproducibility was improved by considering the application conditions of the proposed method.

1. Introduction

3D-free-viewpoint video, which is used to observe a subject from various angles, is actively being researched. Recent 3D-free-viewpoint videos are supposing various situations of a person such as the ballet dance scene [1], the baseball scene [2], the walking scene [3], the golf scene [4] and so on. Images captured synchronously by multiple cameras surrounding a person are used. 3D-free-viewpoint video is generated by re-rendering the appearance of a person from any viewpoint to interactive. In video, the subject can be observed from arbitrary positions that reflect what the observer is interested in. This type of advanced video technology can correctly express spatial relationships between captured subjects, unlike ordinary video that observes using conventional monocular cameras. Industries such as filmmaking and application fields including sports science have increasingly directed their attention to 3D-free-viewpoint video technology. Recent 3D-free-viewpoint videos research, 3D model is constructed using the captured video and rendered to the virtual camera [5,6]. In addition to observing the display image by constructing the virtual environment, it was possible to observe the free viewpoint video using the VR headset [7]. Furthermore, a virtual mirror that can simulate the whole body mirror on the monitor has also been proposed [8]. However, in conventional 3D-free-viewpoint video technology, problems such as the degradation of the quality of the generated video occur because of estimation errors of 3D shapes as well as high calculation costs. Although the effect of estimation errors can be reduced by approximating the object shape to a 2D billboard plane [9], image quality degradation resulting from subject tracking or image segmentation errors remains. Thus, our study examines free-viewpoint technology that utilizes the bullet-time video technique [10] instead of 3D CG modelling and rendering.

One of the advantages of 3D-free-viewpoint video is to let viewers recognize motion parallax, which is a visual effect in which the viewpoint is virtually moving and is very important to feel immersiveness into the scene. By focusing on this point, we research bullet-time video that reproduces this type of visual effect by continuously switching the images captured by multiple cameras that surround the subject [11]. By presenting the captured image nearly as it is, bullet-time video enables us to present high-quality video with little image degradation derived from processing errors. In recent research on bullet-time video, the behavior of a person is the target. A bullet-time video is generated by continuously switching images captured by multiple cameras surrounding a person [12,13]. In our research, a large space like the soccer stadium is the target, and multiple people will be captured. In a large-scale

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: shishido@ccs.tsukuba.ac.jp (H. Shishido), s1520814@u.tsukuba.ac.jp (A. Harazaki), kameda@iit.tsukuba.ac.jp (Y. Kameda), kitahara@ccs.tsukuba.ac.jp (I. Kitahara).

https://doi.org/10.1016/j.jvcir.2019.04.010
1047-3203/

space such as a football stadium, due to the considerable labour and special equipment required for the synchronous capturing of all cameras, sometimes captured multi-view videos are asynchronous. When a bullet-time image is generated using a deviating image, the observation position of the subject as it appears on the image deviates, and the resulting image produces an uncomfortable feeling in the observer. As shown the example in Fig. 1, in a generated bullet-time video, the influence of the synchronization deviation on the observed position of soccer players is small (i.e., the observation position of the players moves in a certain direction on a certain curve). This is because the movement of the player is sufficiently slow relative to the shutter speed of the camera. On the other hand, as shown in Fig. 2 (left), the observation position of the ball moving at a high speed seriously deviates because of the influence of the synchronisation deviation.

Our research aims to reduce the aforementioned deviation on a fast moving object such a soccer ball by using frame interpolation processing, as shown in Fig. 2 (right). There are two contributions of this paper. One is to generate a fine interpolation image in which the objects moving at high speeds are clearly observed (described in Sections 3.2 and 3.3). The other is to choose the appropriate interpolation images to generate a bullet-time video with smooth viewpoint switching. We attempt to reduce the generated discomfort in the viewer by selecting an interpolation image that produces the smoothest change in the observation position of the subject (in Section 3.4).

Our study conducted free-viewpoint technology that utilizes the bullet-time video technique [10]. However, experimental results in free-viewpoint technology [10] are not sufficient to show the effectiveness of the proposed method. Therefore, we prepare some video sequences, in which bullet-time images are switched frame by frame, for both the proposed method and the previous method [11], and carry out subjective/objective evaluation using those video.

In this paper, a subjective evaluation experiment is carried out using bullet-time images generated by the proposed method and the previous method [11], and we verify the effectiveness of the proposed method. We examine the discomfort of the subject in the bullet-time images generated by the proposed method. Furthermore, we clarify how smoothly the proposed method can switch bullet-time images by objective evaluation experiments. And also, reproducibility was improved by considering the application conditions of the proposed method.

In this way, we show the effectiveness of the proposed method by subjective evaluation experiment and objective evaluation experiment.

Fig. 1. Result derived from the comparative task of applying combined processing methods to all camera sequence images. An asynchronous multi-view video is synthesized for the image when the gaze position is that of the player in yellow. The observation position of the player in red moves in a certain direction on an arbitrary curve. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

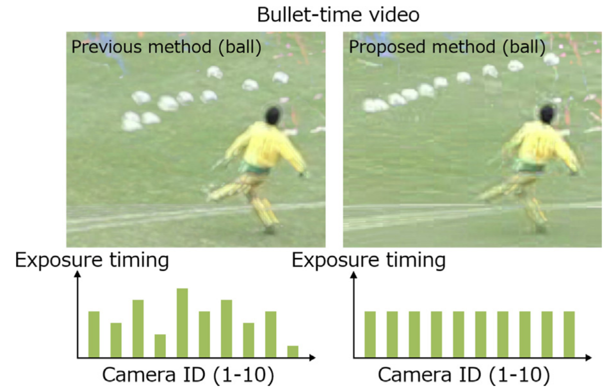


Fig. 2. Result derived from the comparative task that applies combined processing methods to all camera sequence images. Left: When a bullet-time image is captured using multiple asynchronous cameras, the observed position of the ball moving at a high speed deviates because of the effect of the synchronisation deviation. Right: In the proposed method, the deviation is reduced through frame interpolation processing and a smooth viewpoint switching image is generated.

2. Related works

Recently, studies on multi-view video using a large number of cameras targeted at soccer competitions have been reported. Sankoh et al. [14]. proposed an algorithm to realize an immersive free-viewpoint experience in the soccer scene. The homography matrix was well estimated by recognizing correct corresponding feature points between video frames. Kilner et al. [15]. combined simultaneous multi-view shape extraction and stereo refinement. Generation techniques to optimize views of initial scene reconstruction have been demonstrated. Both the shape and appearance of image reconstruction were improved for soccer scene. Jean-Yves et al. [16]. realized high-quality scene representation efficiently by optimization using graph-cut of the energy function combining color, contrast, similarity information and probability in a complex outdoor soccer scene captured on multiple moving cameras at different resolutions. Hilton et al. [17]. used the segmentation and 3D reconstruction of objects from the background using more than four synchronous multi cameras. The playback module rendered the capture scene in real time using a 3D model and texture mapping image computed based on the soccer scene.

In this way, research on multi-view video in a large space such as football stadium is a technology contributing to generation of free viewpoint video. We solve the problem of flickering artifacts (deviations from the observation position) occurring in asynchronous multi-view video of fast-moving objects such as soccer balls. Previous studies have not mentioned this problem. In addition, this research is a technology contributing to the generation of free-viewpoint video by solving this problem.

Some issues remain with respect to the implementation of bullet-time into image capturing technology when actual sports facilities such as soccer stadiums are involved. For example, when we capture multi-view images for 3D computer vision processing or free viewpoint television (FTV) [18], cables for the signal distribution are normally connected to all 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 increases, the establishment costs also increase. In response to such problems, synchronous imaging methods, which can be applied to large-scale spaces, have been proposed. These methods generate external synchronization signals from GPS signals and distribute them to multi-view cameras [19]. However, such special devices increase installation costs. For reasonable image-capturing in practical shooting environments, another method has been proposed that synchronizes multi-view asynchronous images using audio information [20]. However, in our target environment, such as a soccer stadium, critical echo effects (i.e., reflection sound) complicate the synchronization process. If a system can achieve frame-level synchronization, shutter-level asynchronicity remains because the internal clocks of the cameras are not synchronized.

As previously mentioned, if we consider accurate 3D computer vision processing, completely synchronized multi-view videos are required that consume a huge amount of labour and use special equipment. By contrast, a bullet-time video can be generated using asynchronous videos because it does not require accurate 3D information. However, when we capture fast-moving objects, flickering artefacts (deviations from the observed position) are observed in the bullet-time video.

In this research, we reduce this deviation using a frame interpolation technique for generating bullet-time videos that smoothly switches the viewpoint. In each multi-view video, frame interpolation processing is applied to the frames before and after the frame of interest. Frame interpolation processing is generally achieved through image morphing in which a series of image groups are generated whose appearance changes continuously from one shape to another. Frame interpolation uses a process that synthesizes the appearance of a middle part between two images [21]. In a typical morphing method [22], the correspondence information between two input images is manually given. However, when video data consisting of 30 or more frames per second (fps) are processed, manually inputting information that requires correspondence is not feasible. To solve this problem, we employ regenerative morphing [23], 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 considerably different. As a result, generating interpolation images in which the object is clearly observed at the desired position is difficult. In this research, we address this problem by generating an initial interpolation image that includes the visual information cue in order to generate an appropriate interpolation image from a simple correspondence.

3. Smooth multi-view image switching processing

In this section, we outline the smooth multi-view image switching process. The proposed method consists of two main processes.

The first is to generate multiple interpolation images between consecutive front and rear frames using frame interpolation processing at each view-point, as shown in Fig. 3 ①. In the proposed method, the front and rear frames are targeted in each camera. Multiple interpolation images are generated using the morphing method, where the latter adopts regenerative morphing. In such cases, for objects moving at a high speed, initial motion blur interpolation images are used with simple correspondence information as clues. Therefore, an interpolation image in which the subject is observed at an appropriate position is generated.

In the second case, as shown in Fig. 3 ②, an image with little change in the observation position of the subject out of the generated interpolation images is selected. This generates a bullet-time image in which viewer discomfort at switching is reduced. An image having the smallest deviation in the observation position of the subject moving at a high speed from the generated interpolation image is selected. The image surrounded by a pink colour frame is used as a switching image.

In Section 3.1, we describe the problems of Regenerative Morphing. In Sections 3.2 and 3.3, we describe a frame interpolation method using regenerative morphing. In Section 3.4, we describe the method of selecting the generated interpolated image.

3.1. Usefulness and problems of regenerative morphing

We applied regenerative morphing for frame interpolation processing. Fig. 4 shows the interpolated image T_n between input images S_1 and S_2. Fig. 4(a) shows image generation when blending S_1 and S_2. In target interpolation image T_n, interpolation image generation is realized as an optimization problem of the algorithm of [24]. The algorithm of [24] optimizes the bidirectional dissimilarity (error) measure using the consistency between the adjacent and previous interpolation images and the similarity of the input images. Fig. 4(b) is an interpolated image generated by the initial image. Here, the blended image (Fig. 4(a), middle) is used as the interpolation's initial image.

In the initial image for the interpolation, because no visual cue information of the target object is present around the images centre where the target object should be observed, in the interpolation result (Fig. 4(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 the information of the target object's appearance does not exist at an appropriate position in the initial interpolation image.

3.2. Frame interpolation using regenerative morphing

We solve this problem by including just one set of correspondence information about the subjects between the input images. Just one set of correspondence information is the centre coordinates of the subject shown in Fig. 4. That is, the centre coordinates $S_1(x_1, y_1)$ of the ball at S_1 and the centre coordinates $S_2(x_2, y_2)$ of the ball at S_2 are taken as a one set of correspondence information. This

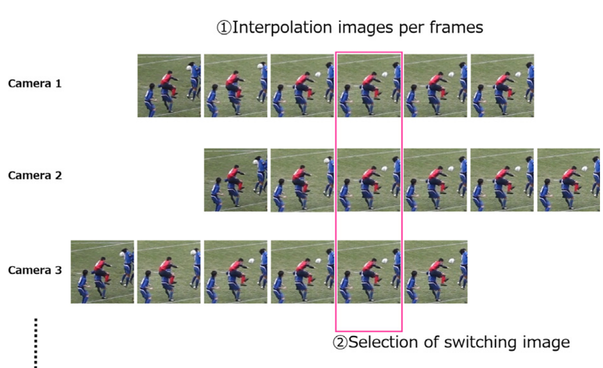


Fig. 3. Schematic of the proposed method.

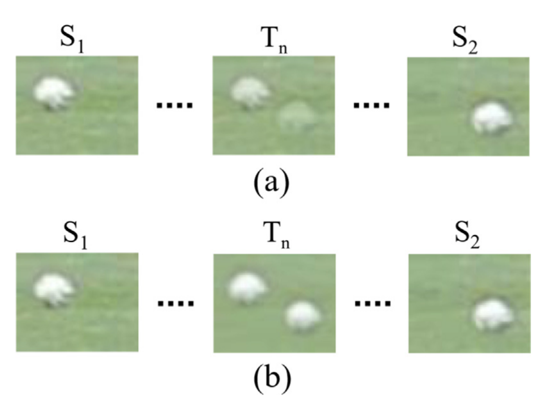


Fig. 4. (a) Image generation when blending S_1 and S_2, (b) Interpolated image generated by the initial image.

coordinate value is acquired using the ball tracking method [25]. 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. 4 (a), middle). The motion blur area that is added depends on the target object's size. As shown in Fig. 5, 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).

$$H : h\frac{x}{X} = D : d \tag{1}$$

$$h = Hd \times \frac{X}{Dx} \tag{2}$$

3.3. Determination of blend ratio between subject and background

To generate a plurality of interpolation images with different interpolation distances between input images, we determined the blending ratio of the subject and background based on the interpolation distance between the two input images using the normal distribution shown in the following Eq. (3). As shown in Fig. 6, let t be the position between the input images, T_n be the ratio of the interpolation distance to the input image, and σ^2 be the variance value.

$$f_{T_n}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t - T_n)^2}{2\sigma^2}\right) \tag{3}$$

The larger the value of the expression is, the larger the information of the subject. The smaller the value of the expression is, the larger the background information becomes. In this way, we calculate the blending of the subject and the background according to the interpolation distance. Fig. 7(a) shows the initial interpolation image after applying the motion blur effect. Fig. 7(b) shows an interpolation image generated using the initial interpolation images. As shown in Fig. 7(b), it can be confirmed by the proposed method that the target object is observed as a single image at an appropriate position (in this example, in the middle of the ball observation position of S_1 and S_2). By applying this generated image to bullet-time video, smooth viewpoint switching is realized.

3.4. Selection method of generated interpolation image

We next describe a method for choosing an appropriate interpolation image that smoothly changes the observation position of the target object when the viewpoint is moving.

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 the angle θ_n between

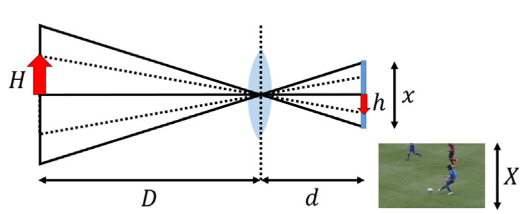


Fig. 5. Size of the subject on the image.

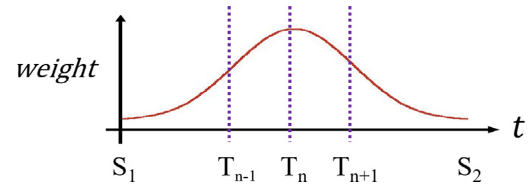


Fig. 6. The blend ratio of the subject and the background is determined based on the interpolation distance between the two input images using the normal distribution.

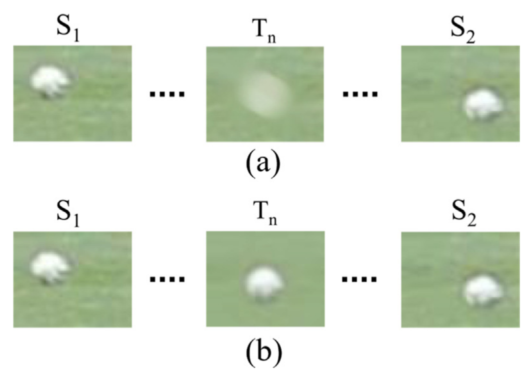


Fig. 7. (a) Initial interpolation image after applying the motion blur effect. (b) Interpolation image generated using the initial interpolation images.

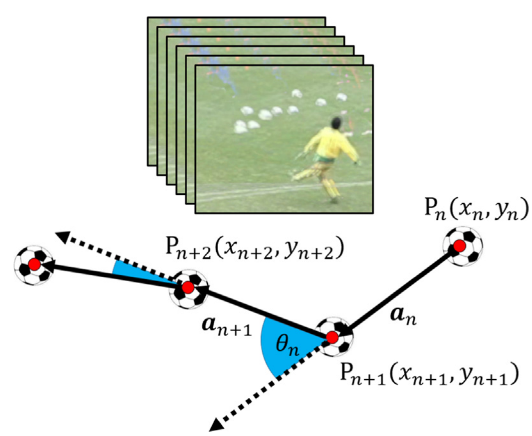


Fig. 8. In all of the input and interpolation images, we acquired the observation position of the object. Next, we calculated the angle θ_n between the following two vectors.

the following two vectors (Fig. 8). Vector a_n is 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 a_{n+1} is 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 accumulated components of this angle were calculated by all the cameras (Cam1, ... ,CamN) using the following Eq. (4). Similar processing was performed for all the combinations of input and interpolation images of the multi-view cameras, and an image was chosen that had the minimum accumulation values. The smaller the angle between the two vectors was, the greater the number of coordinates of the three points that were in the constant curve shape. The lower the resulting value of (4) was, the smaller the change in the observation position of the object. Finally, as shown in Fig. 3, smooth image switching was then performed.

$$\sum_{n=1}^{N-1}\theta_n = \sum_{n=1}^{N-1}cos^{-1}\left(\frac{a_n \cdot a_{n+1}}{|a_n||a_{n+1}|}\right) \quad (4)$$

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

4. Experiment and results

4.1. Experiment with a large soccer stadium

We conducted a video-capturing experiment based on a youth soccer match played in a large soccer stadium. As shown in Fig. 9, we installed 10 video cameras (Canon EOS 5D Mark II, 1920 pixels × 1080 pixels at 30 fps) in the seats behind the goal to capture the near-penalty area. The convergence angle of the neighbouring cameras was approximately 10 degrees. Image processing was performed using a computer with the following specifications: CPU: Intel Core i7 3.4 GHz, GPU: Intel HD Graphics Family, and RAM: 8.0 G.

We confirmed the effectiveness of the proposed method for another sequence image. Fig. 10 show a bullet-time image obtained when switching the multi-view images, and Fig. 10 show a bullet-time image obtained from using our proposed method. The observation position of the soccer ball in a point changes from yellow to red.

When we switch the image, the soccer ball's position moves forward and backward (Fig. 10(a), (c), (e)). However, in Fig. 10(b), (d) and (f), it moves in a constant direction along a smooth curve.

In particular, in Fig. 10(a), the trajectory of the observation position greatly deviates, but that problem was eliminated in Fig. 10(b). By minimizing the sum of the angles obtained from the three observation coordinates, all of the observed coordinates were corrected. This step resulted in smoother switching.

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

4.2. Subjective evaluation experiment using bullet-time video

A subjective evaluation experiment was carried out using bullet-time images generated by the proposed method and the previous method [11], and we verified the effectiveness of the proposed method. Subjects were allowed to observe the images generated by the proposed method and the previous method [11] once.

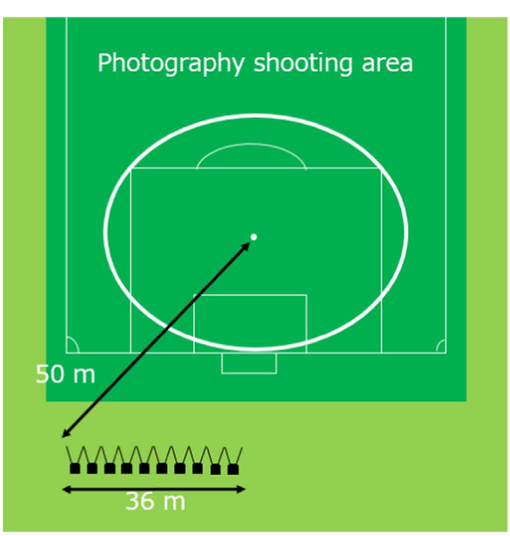


Fig. 9. Image capturing the environment's outline.

After that, we instructed them to select the image with less discomfort. To cancel the learning effect, we set the order of showing the two images randomly. The subjects were 12 people (11 males and 1 female) aged 22 to 29.

The videos used for the subjective evaluation experiment had the following 4 types (8 videos).

[Video playback speed (fast): 265 kb/s]

Video A: Video set with a yellow player as the gaze point (proposed method: Fig. 10(a), previous method [11]: Fig. 10(b)).

Video B: Video set with a blue player as the gaze point (proposed method: Fig. 10(c), previous method [11]: Fig. 10(d)).

[Video playback speed (slow): 141 kb/s]

Video C: Video set with a yellow player as the gaze point (proposed method: Fig. 10(a), previous method [11]: Fig. 10(b)).

Video D: Video set with a blue player as the gaze point (proposed method: Fig. 10(c), previous method [11]: Fig. 10(d)).

The experimental results are shown in Fig. 11. On the left of Fig. 11 is the experimental result for the image in which the yellow player is set as the gaze point. On the right of Fig. 11 is the experimental result for the image in which the blue player is set as the gaze point.

The horizontal axis represents three types of answer items, including the "Proposed method", the "Previous method", and "Neither". The vertical axis shows the number of people answering each item. The purple bars represent the results of videos played at a fast speed. The cyan bars show the results of the videos played at a slow speed.

Next, a 2-sample t-test was conducted. The level of incompatibility was "no incompatibility: 2", "neither: 1", and "incompatibility: 0".

As a result of the 2-sample t-test, the p-value of the experiment for video A was 0.0004, thus indicating a significant difference. Similarly, the p-value of the experiment for video B was 0.0000004, thus indicating a significant difference. As a result, in the experiments using two types of video with a fast video playback speed (265 kb/s), it became clear that the subjects had less discomfort in the bullet-time video generated by the proposed method.

The p-value of the experiment for video C was 0.006, thus indicating a significant difference. Similarly, the p-value of the experiment for the video D was 0.654, thus indicating no significant difference. As a result, it was revealed that in video C, the subjects had less discomfort in the bullet-time video generated by the proposed method. Since a significant difference was not indicated in video D, in the case of the bullet-time video where the video playback speed was slow, the influence on the video quality was reduced. On the other hand, when the playback speed is fast, it can be considered that the influence on the video quality of the bullet-time video increases.

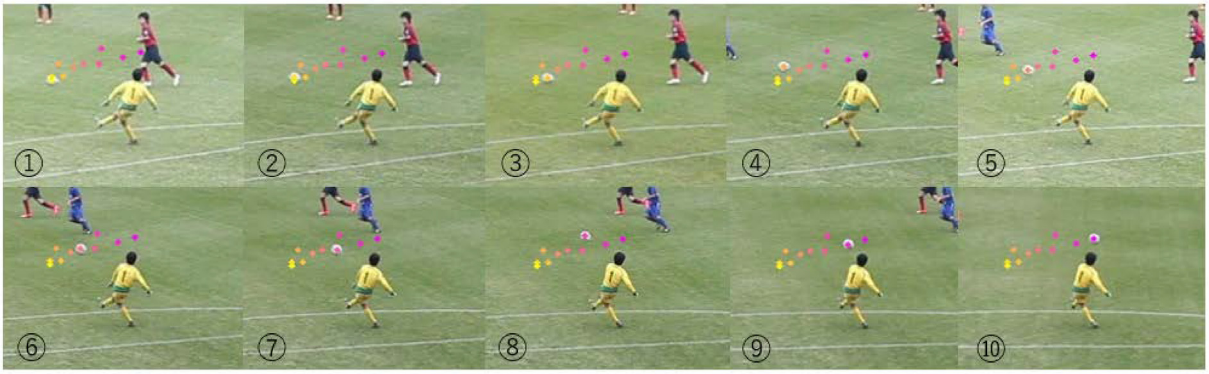
4.3. Objective evaluation experiment using bullet-time video

We conducted objective evaluation experiments using 3 types and 6 videos.

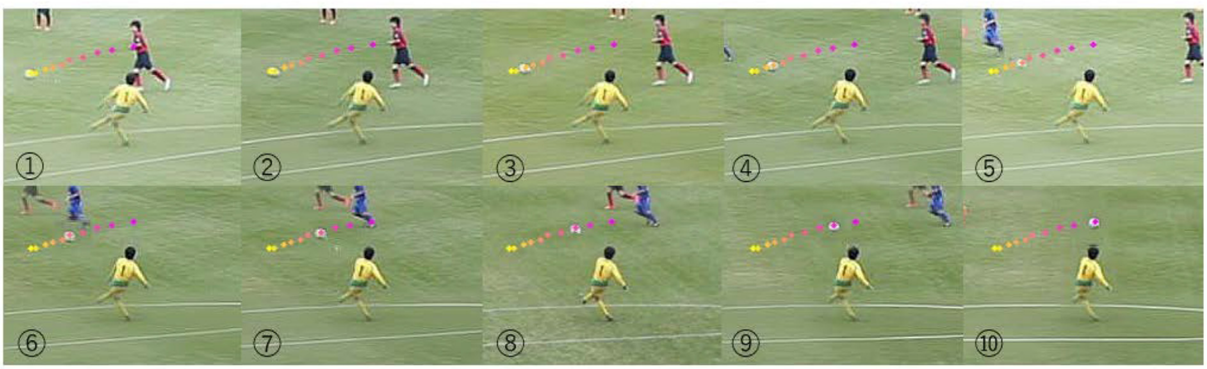
- Video set with a yellow player as the gaze point (previous method [11]: Fig. 10(a), proposed method: Fig. 10(b))
- Video set with a blue player as the gaze point (previous method [11]: Fig. 10(c), proposed method: Fig. 10(d))
- Video set with a red player as the gaze point (previous method [11]: Fig. 10(e), proposed method: Fig. 10(f))

We manually calculate the centre coordinates of the balls in three frames from the generated bullet-time video. The angle θ of the previous method [11] and the proposed method is calculated using the Eq. (4, 5). The results are shown in the Table 1.

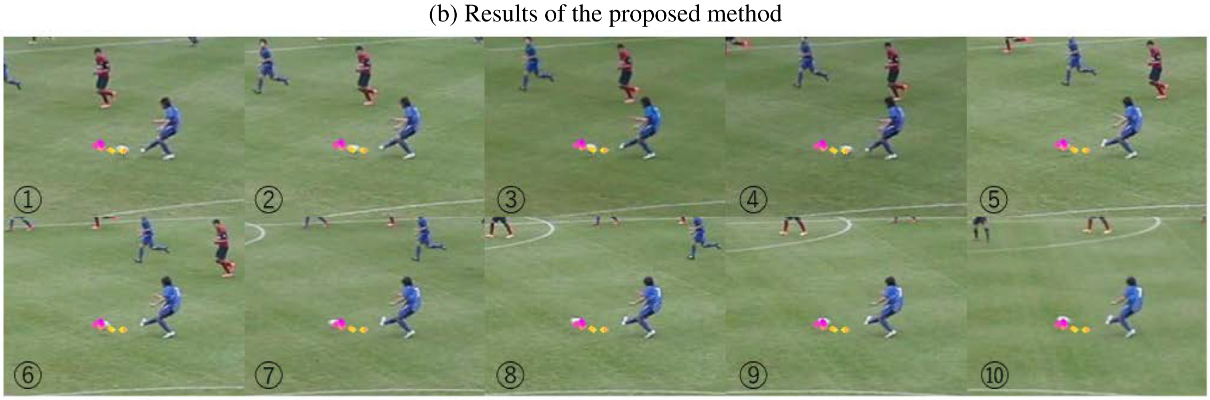
For the three types of videos, we can confirm that the angle θ of the proposed method is smaller than the previous method [11]. Similarly, the value of SD can be confirmed to be smaller than the previous method [11]. The smaller that the value of Eq. (4) is, the smaller that the observation position of the subject changes and the smoother the image switching. Therefore, it can be



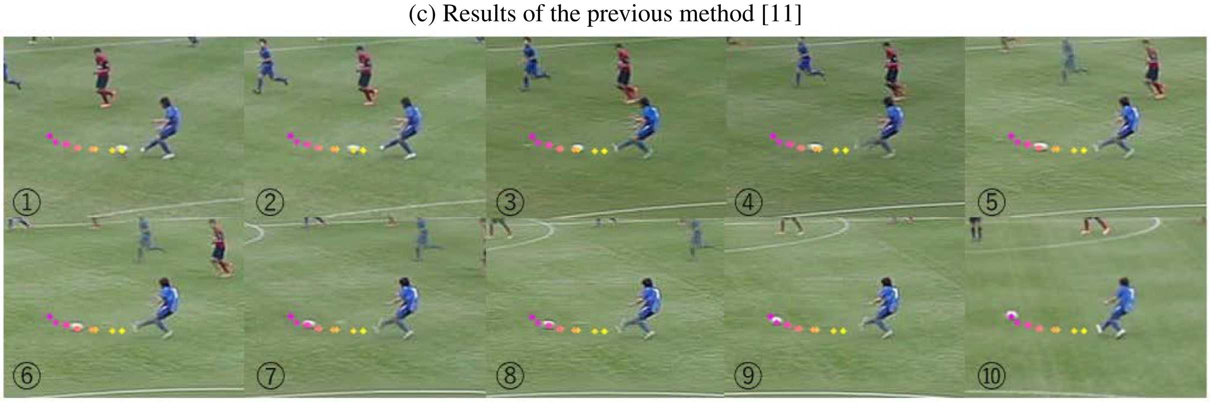
(a) Results of the previous method [11]



(b) Results of the proposed method

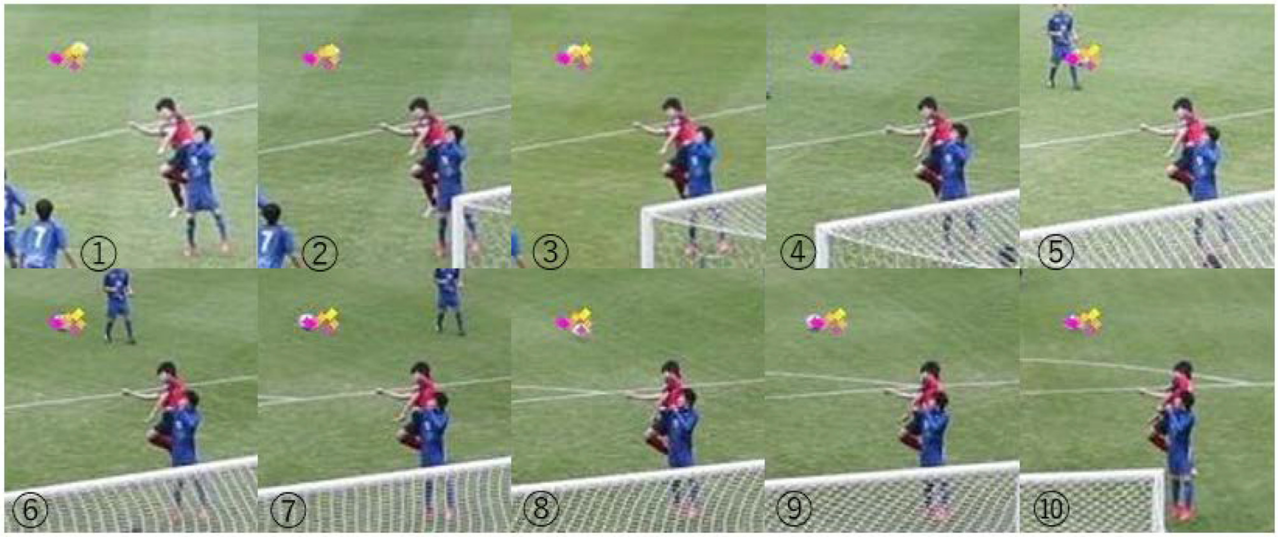


(c) Results of the previous method [11]

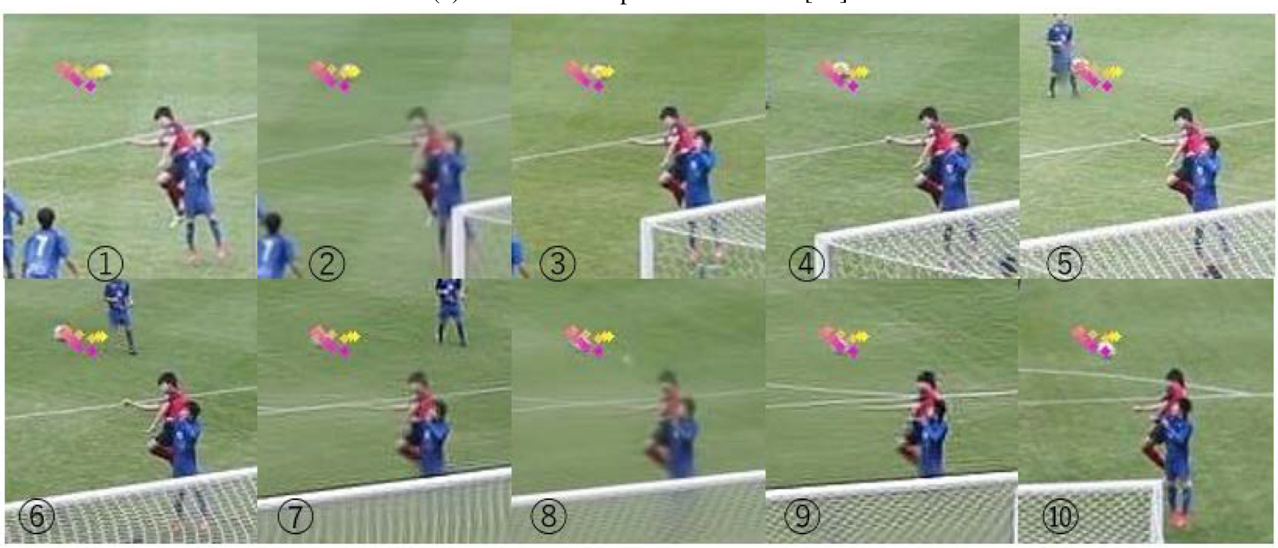


(d) Results of the proposed method

Fig. 10. Resulting images obtained when applying the proposed method: The number at the lower left indicates the camera number and the order in which images were switched.

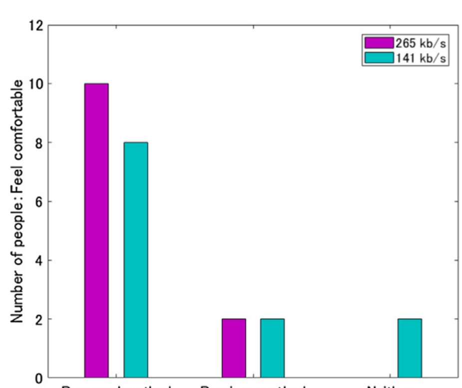


(e) Results of the previous method [11]

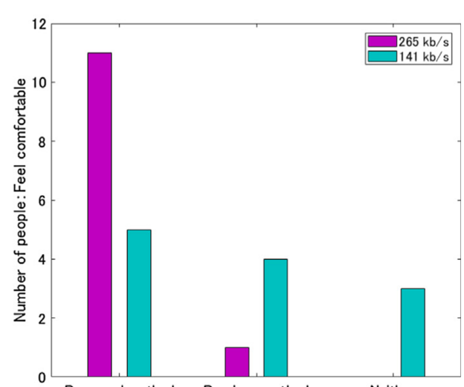


(f) Results of the proposed method

Fig. 10 (*continued*)



Yellow players video



Blue players video

Fig. 11. Subjective evaluation experiment using bullet-time video generated by the proposed method and the previous method [11].

Table 1
The angle θ of the previous method [11] and the proposed method is calculated using the Eqs. (4) and (5).

	Frame	1,2,3	2,3,4	3,4,5	4,5,6	5,6,7	6,7,8	7,8,9	8,9,10	Avg.	SD
Yellow player	Proposed (Degree)	20.46	2.03	4.31	15.62	18.34	5.70	7.55	0.45	9.31	7.23
	Previous (Degree)	89.91	68.20	54.16	34.50	20.46	75.87	84.60	31.44	57.39	24.56
Blue player	Proposed (Degree)	3.99	14.26	12.63	32.18	47.03	20.46	20.46	24.35	21.92	12.31
	Previous (Degree)	7.03	4.40	1.31	32.28	34.70	53.13	44.91	89.91	33.46	28.03
Red player	Proposed (Degree)	18.34	36.78	0.00	44.91	108.34	11.22	11.22	26.47	32.16	31.86
	Previous (Degree)	30.87	139.31	160.47	7.03	0.00	158.11	179.91	175.15	106.36	73.92

considered that the proposed method produces a smoothly switched image rather than the previous method [11]. In this way, the effectiveness of the proposed method was confirmed in the scene of Fig. 10. The videos used in the evaluation experiment are three types, the goalkeeper kicks the ball, the player gives a pass, the player shoots with a header. The ball speed at this time is different. We confirmed that the proposed method can be applied to various speed ranges of fast moving objects (soccer balls). Next, as shown in Fig. 12, a comparison experiment of the previous interpolation image generation method and the proposed method was performed. Regenerative Morphing [23] was used for the previous method to be compared. The input image used for the experiment was a soccer ball image of the front and rear frames of the multi-view video (Camera 4, 5, 6). As shown in the lower part of Fig. 12, two soccer balls are generated as interpolation images of the previous method [23]. This is because there is no subject clue at an appropriate position on the initial interpolation image when the subject moves at high speed. A simple blending result is set as an initial interpolation image in the previous method [23]. On the other hand, as shown in the upper part of Fig. 12, only one soccer ball is generated as an interpolation image of the proposed method. Furthermore, the soccer ball position is gradually moving downward in Camera 4, 5 and 6. An image in which a soccer ball is moving with the movement of the camera can be generated. Therefore, the interpolation image generated by the proposed method has no flickering artefacts. In addition, we can confirm the effectiveness of the initial interpolation image due to the weight of motion blur.

4.4. Conditions for applying objects

We consider the conditions for applying objects moving at high speed in the frame interpolation process with the proposed method. Let S be the diameter of the soccer ball, as shown in Fig. 13(a). The distance between the ball centres in the two frames is d, as shown in Fig. 13(b)-(d). The interpolation image is generated by satisfying conditions (b) and (c) using the regenerative morphing method [23]. Therefore, when $d \leqq S$, the frame interpolation image can be generated with high quality. Conversely, as d becomes larger than S, the quality of the frame interpolation image deteriorates, as shown in Fig. 13(d). In general, the passing speed of a soccer ball is V, which is approximately 25 m/s [26]. The ball size

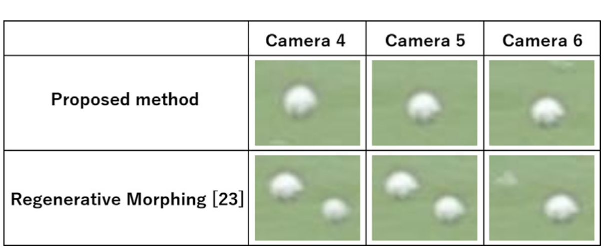


Fig. 12. Comparison of previous interpolation image generation method [23] and proposed method.

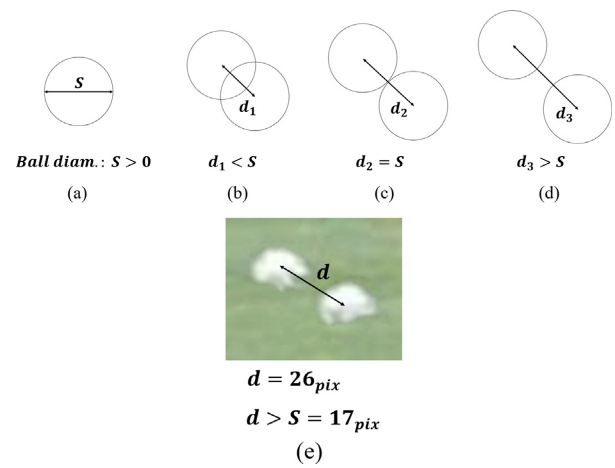


Fig. 13. (a) Soccer ball diameter: S, (b-d) The distance between the ball centres in the two frames: d, and (e) The distance d of the experimental environment.

(diameter) is $S = 22$ cm. The ball diameter in this experimental environment was $s = 17$ pixels. The ball velocity on the image can be calculated as $v = 19$ pixels/s. d was 26 pixels in this experiment, as shown in Fig. 13(e). Therefore, when $d > S$, if the distance is at least $S + 9$ pixels, an acceptable frame interpolation image can be generated in this experimental environment.

5. Conclusions

This research proposed a method that generates viewpoint smooth switching by reducing the flickering artefacts that are observed at bullet-times generated from asynchronous multi-view videos using frame interpolation processing.

In the frame interpolation processing, previous methods (Regenerative morphing) failed to deal with flickering artifacts observed in asynchronous multi-view videos. In the proposed method, the problem was solved by generating an initial interpolation image based on the weight of motion blur. We conducted a comparative experiment between the previous method and the proposed method and showed the effectiveness of the proposed method. Also, regarding the problem of positional shift of the subject caused by the asynchronous image, smooth image generation was realized by calculating the angle between the vectors of the subjects in the continuous frame.

A subjective evaluation experiment was carried out using bullet-time images generated by the proposed method and the previous method, and we verified the effectiveness of the proposed method. We examined the discomfort of the subject in the bullet-time images generated by the proposed method. An objective evaluation experiments was carried out using various videos about generation of smooth video. And we confirmed that the proposed method can be applied to various speed ranges of fast moving objects (soccer balls). In addition, reproducibility was improved by considering the application conditions of the proposed method.

In this way, we showed the effectiveness of the proposed method by subjective evaluation experiment and objective evaluation experiment.

In order to further improve the quality of the result, the resolution must be set high. However, processing cost is high if the resolution is increased. Therefore, we must devise a method to process the proposed method at high speed. If this problem can be solved, it will be applicable to current TV broadcasting.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Nos. 17H01772 and 17K13180 and CREST, JST.

References

[1] J. Carranza, C. Theobalt, M. Magnor, H. Seidel, Free-viewpoint video of human actors, ACM Trans. Graph. 22 (2003) 569–577.
[2] T. Kanade, P. Rander, P. Narayanan, Virtualizedreality: Constructing virtual worlds from real scenes, IEEE Multimedia, Immers. Telepres. 4 (1997) 34–47.
[3] Y. Liu, Q. Dai, W. Xu, A point-cloud-based multi-view stereo algorithm for free-viewpoint video, IEEE Trans. Visual Comput. Graphics 16 (2009) 407–418.
[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 Trans. Graph. 34 (2015).
[5] J. Silva, T. Santos, C. Morimoto, Automatic camera control in virtual environments augmented using multiple sparse videos, Comput. Graph. 35 (2011) 412–421.
[6] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, 3d video and free viewpoint video - technologies, in: Applications and MPEG Standards, 2006 IEEE International Conference on Multimedia and Expo, 2006, pp. 2161–2164.
[7] N. O'Dwyer, N. Johnson, E. Bates, R. Pages, J. Ondrej, K. Amplianitis, D. Monaghan, A. Smolic, Virtual play in free-viewpoint video: Reinterpreting samuel beckett for virtual reality, in: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), 2017, pp. 262–267.
[8] M. Straka, S. Hauswiesner, M. Ruther, H. Bischof, A free-viewpoint virtual mirror with marker-less user interaction, in: Scandinavian Conference on Image Analysis (SCIA), 2011, pp. 635–645.
[9] T. Koyama, I. Kitahata, Y. Ohta, Live mixed- reality 3d video in soccer stadium, in: 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, 2003, pp. 178–186.
[10] H. Shishido, A. Harazaki, Y. Kameda, I. Kitahara, Smoothly switching method of asynchronous multi-view videos using frame interpolation, in: 3DTV Conference 2017: Research and Applications in Future 3D (3DTV-CON), 2017, pp. 1–4.
[11] N. Akechi, I. Kitahara, R. Sakamoto, Y. Ohta, Multi-resolution bullet-time effect, in: SIGGRAPH-ASIA2014, vol. 30, 2014.
[12] A. Smolic, 3d video and free viewpoint video-from capture to display, Pattern Recogn. 44 (2011) 1958–1968.
[13] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High-quality video view interpolation using a layered representation, in: ACM SIGGRAPH, 2004, pp. 600–608.
[14] H. Sankoh, S. Naito, M. Harada, T. Sakata, M. Minoh, Free-viewpoint video synthesis for sport scenes captured with a single moving camera, ITE Trans. Media Technol. Appl. (MTA) 3 (2015) 48–57.
[15] J. Kilner, J. Starck, A. Hilton, O. Grau, Dual-mode deformable models for free-viewpoint video of sports events, in: Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM, 2007), 2007, pp. 177–184.
[16] J. Guillemaut, J. Kilner, A. Hilton, Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 809–816.
[17] A. Hilton, J. Guillemaut, J. Kilner, O. Grau, G. Thomas, Free-viewpoint video for tv sport production, Image Geomet. Process. 3-D Cinematograp. 5 (2010) 77–106.
[18] M. Tanimoto, M. Tehrani, T. Fujii, Free-viewpoint tv, IEEE Signal Process. Mag. 28 (2011) 67–76.
[19] I. Kitahara, H. Saito, Y. Ohta, T. Ono, T. Kanade, Large-scale virtualized reality, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Technical Sketches, 2001.
[20] RedGiant, Plural eyes, in: https://www.redgiant.com/products/pluraleyes/, 2014.
[21] T. Beier, S. Neely, Feature-based image metamorphosis, in: ACM SIGGRAPH Computer Graphics, 1992, pp. 35–42.
[22] S. Lee, K. Chwa, J. Hahn, S. Shin, Image morphing using deformation techniques, J. Visual. Comput. Animat. 7 (1996) 3–23.
[23] E. Shechtman, A. Rav-Acha, M. Irani, S. Seitz, Regenerative morphing, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 615–622.
[24] D. Simakov, Y. Caspi, E. Shechtman, M. Irani, Summarizing visual data using bidirectional similarity, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[25] K. Choi, Y. Seo, Tracking soccer ball in tv broadcast video, in: Image Analysis and Processing - ICIAP 2005, 2005, pp. 661–668.
[26] E. Kellis, A. Katis, Biomechanical characteristics and determinants of instep soccer kick, J. Sports Sci. Med. 6 (2007) 154–165.