Pseudo-Dolly-In Video Generation Combining 3D Modeling and Image Reconstruction

Hidehiko Shishido*   Kazuki Yamanaka [†]   Yoshinari Kameda [‡]   Itaru Kitahara [§]

University of Tsukuba

Figure 1: Images of a soccer game. Upper part shows pseudo-dolly video by proposed method. In the pseudo-dolly-in video, motion parallax can be reproduced and the depth can be understood. Furthermore, since the reconstruction process is applied to the missing region, the proposed method maintains the image quality. Lower figure shows digital zoom image. When comparing with digital zoom image and proposed method, in the proposed method, the players are rendered at (relatively) geometrically correct positions as if the camera goes in to the field (dolly-in).

ABSTRACT

This paper proposes a pseudo-dolly-in video generation method that reproduces motion parallax by applying image reconstruction processing to multi-view videos. Since dolly-in video is taken by moving a camera forward to reproduce motion parallax, we can present a sense of immersion. However, at a sporting event in a large-scale space, moving a camera is difficult. Our research generates dolly-in video from multi-view images captured by fixed cameras. By applying the Image-Based Modeling technique, dolly-in video can be generated. Unfortunately, the video quality is often damaged by the 3D estimation error. On the other hand, Bullet-Time realizes high-quality video observation. However, moving the virtual-viewpoint from the capturing positions is difficult. To solve these problems, we propose a method to generate a pseudo-dolly-in image by installing 3D estimation and image reconstruction techniques into Bullet-Time and show its effectiveness by applying it to multi-view videos captured at an actual soccer stadium. In the experiment, we compared the proposed method with digital zoom images and with the dolly-in video generated from the Image-Based Modeling and Rendering method.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

---

* e-mail: shishido@ccs.tsukuba.ac.jp

† e-mail: s1520831@u.tsukuba.ac.jp

‡ e-mail: kameda@iit.tsukuba.ac.jp

§ e-mail: kitahara@ccs.tsukuba.ac.jp

1 INTRODUCTION

In this paper, we propose a method for generating pseudo-dolly-in video that can reproduce motion parallax based on 3D estimation from multi-view videos (figure 1). To improve the generated video quality, we apply image reconstruction processing. Dolly is a video-shooting technique that places a camera on a trolley (figure 2) and captures video by moving the trolley on a rail. Dolly-in video, which is captured by moving the trolley toward the subject, can reproduce motion parallax, allowing viewers to perceive depth and immersive sensations. Therefore, this technique can express the spread of space in television dramas and movies. However, for sporting events in large-scale spaces, it is difficult to set a rail that moves the trolley required for capturing dolly-in video at stadiums. In this research, we propose video generation that produces visual effects that are equivalent to dolly-in video using multi-view images captured with multiple cameras fixed outside the field.

The Image-Based Modeling and Rendering (IBMR) technique [1-10] reproduces appearances from an arbitrary viewpoint by analyzing multi-view images. By reconstructing the subject's 3D model, we can reproduce its appearance from a viewpoint at which we cannot actually place the camera. In the same way, dolly-in video can also be generated. However, due to such influences as the calibration error of multi-view camera, synchronization shift, changes in lighting conditions, and appearance changes due to sunlight fluctuation, the generated 3D model often includes estimation error [11]. Therefore, [11] has a problem that the quality of the generated dolly-in video may deteriorate.

Bullet-Time, which is another way to reproduce the viewpoint movement sensation, sequentially switches the multi-view images captured by fixed multiple cameras [12,13,14]. Although it maintains the quality of the captured video, unfortunately, the viewpoint's movable area is limited to where the shooting camera is installed. Therefore, for observing the detailed parts of a target object, we improved the Bullet-Time effect by installing a digital zooming function [12] that makes it possible to observe detailed

IEEE computer society

information even when images are captured at locations away from the object, such as large-scale, outdoor spaces. However, although the observation size of the target object is changed by the digital zoom, when the shooting camera is moved back and forth, the motion parallax like dolly-in video is not reproduced by the digital zoom. Therefore, providing immersive sensations is difficult [15].

We propose a method that generates pseudo-dolly-in video by applying 3D estimation and image reconstruction to multi-view images captured from multiple fixed cameras. The proposed method is outlined in figure 2. Instead of directly generating a dolly-in image using IBMR technology, the depth information of the target scene photographed by a virtual dolly-in camera is generated from an estimated 3D model. After mapping the texture from multiple images to the dolly-in image by referring to the depth information, we apply image reconstruction processing [16] for hole-filling to generate high-quality dolly-in video.
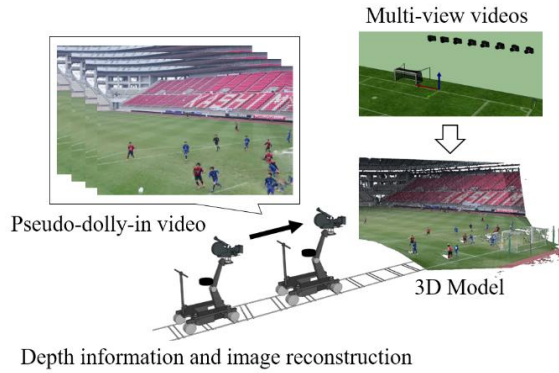

Figure 2: Pseudo-dolly-in video generation using depth information obtained from 3D model and image reconstruction processing. Rather than directly acquiring dolly-in-images from 3D model restored using multi-view image, reconstruction processing is done on it with depth information generated from 3D model as a clue. Then a pseudo-dolly-in picture is generated.

2 RELATED WORK

2.1 Dolly-in and zoom-in video

Dolly-in video separates objects from their background, but there is a difference of the object and the background mix in zoomed-in video [17]. The left column of figure 3 shows digital zoom-in video that enlarged the target. The right column shows dolly-in video captured by a moving camera. When comparing these videos, the dolly-in video gives a perception of the subject's surrounding space; the digital zoom-in video fails to understand its surrounding space.

In recent years, a system has been developed that utilizes the dolly-in effect [18]. To enhance the reality of video phones, the system physically moves a shooting camera on its other side along the optical axis. The camera moves back and forth in accordance with the movement of the viewpoint of the other side user. In evaluative experiments using the developed system, we confirmed that video, in which the motion parallax in the back and forth direction is reproduced, gives a higher immersive sensation. In such research, we can reproduce a wider region of motion parallax by an RGB-D camera. However, in a large-scale, outdoor space, as targeted in our research, the capturing distance becomes too long to accurately acquire depth information. Therefore, we estimate the depth information by calculating the 3D information of the target space from multi-view videos to generate dolly-in video.
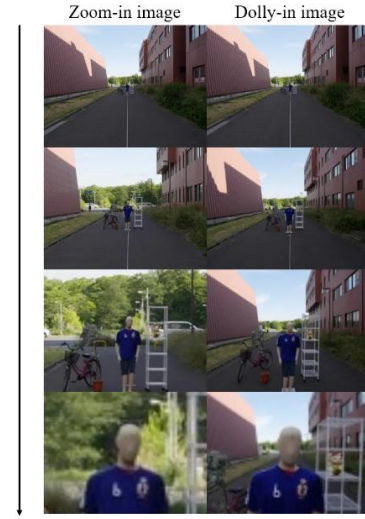

Figure 3: Image captured from distant object: Left: zoom-in image of object; Right: dolly-in image of object. Dolly-in video gives sense of the spread of its surrounding space.

2.2 IBMR

IBMR research has been actively conducted [1-10]. The Shape from Silhouette (SfS) method, which is one popular method for estimating 3D models, estimates the 3D information of a target by merging the silhouette regions extracted in each multi-image. As the number of multiple cameras increases, the 3D estimation accuracy is also improved as well as the quality of the generated free-viewpoint video. However, SfS estimation accuracy is easily damaged by the noise of the captured images [11].

2.3 Generating Bullet-Time video

The Bullet-Time method presents the viewpoint movement sensation by continuously switching the multi-view video based on camera arrangement. Since the captured image is presented almost live movie, one of its features is that the degradation of the quality due to image processing is relatively small.

To reproduce the visual effect of viewpoint movement in Bullet-Time, all of the cameras must be installed so that the optical axis of the multi-view camera crosses at one point in the image space. The Eye Vision [20] camera control system has a robotic arm to achieve the above conditions in dynamic scenes. However, when these systems capture images, the operators must manually input the points to be noticed, and in some situations such points of interest cannot be set because it is impossible to predict the object's motion, like in sporting events. Therefore, we developed a Bullet-Time system in which the image viewer can set the gaze point manually in the image of the fixed multi-view camera. The viewer can move the viewpoint without missing an object [12]. However, since Bullet-Time is a switching presentation of 2D images, it is difficult to move the virtual camera into the target space (i.e., an immersive viewpoint), as in the IBMR method.

In this research, we propose a method that generates high-quality Bullet-Time video that can reproduce motion parallax by combining 3D-shape and image reconstruction processing.

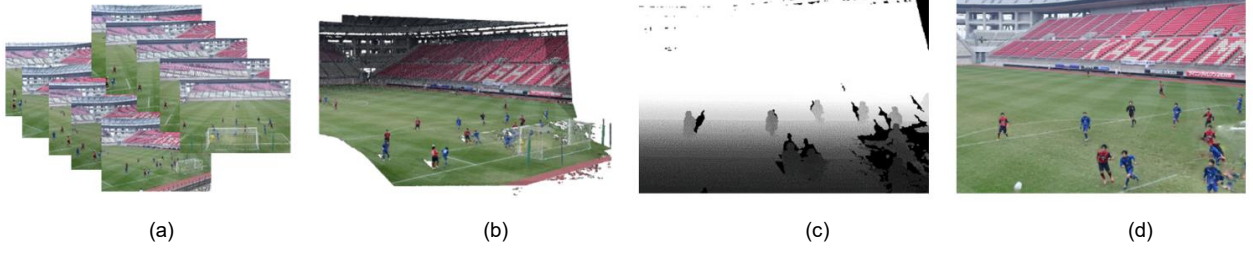(a)   (b)   (c)   (d)

Figure 4: Flow of proposed method:

(a) multi-view videos; (b) restoration of 3D model; (c) generation of depth map; (d) generation of pseudo-dolly-in video

2.4   Image reconstruction processing

Image reconstruction processing generates an image to avoid impairing such important regions as the interested objects in the image as much as possible and to avoid disturbing the visual continuity [16]. Seam Carving [21], which is a well developed image reconstruction processing, removes/enlarges pixels for pixels that are considered unimportant as image content based on gradient information. [21] realized Image resizing processing that maintains visual continuity. Furthermore, by considering the depth information acquired from RGB-D cameras, the detection accuracy of important and non-critical regions can be improved, including better visual continuity of generated images [22]. We utilize this process for high-quality image generation that maintains visual continuity.

3   PSEUDO-DOLLY-IN IMAGE GENERATION USING IMAGE RECONSTRUCTION PROCESSING

In our research, multi-view images (figure 4(a)) are captured by multiple cameras fixed in such large-scale outdoor spaces as a soccer stadium. Multi-view cameras are calibrated by referring to corresponding points among the multi-view videos (figure 4(b)). 3D-point clouds are also generated, and a 3D model is estimated based on them. In the 3D model, we put a virtual camera (a dolly-in camera) and calculate the position and orientation. By projecting the 3D model onto the image captured using the camera parameter of the dolly-in camera, a depth map of the dolly-in camera is generated (figure 4(c)).

We mapped the texture of a multi-view image to the dolly-in image using depth map (figure 4(d)). However, it is difficult to reproduce the appearance of regions that are not visible in multi-view images (black in the depth image, figure 4(c)). So we compensate for the regions with other regions of this image with a similar appearance to the target region by image reconstruction processing. The filling process keeps the image's visual continuity to generate a high quality image.

Figure 5: Left: point clouds restored by SfM; Right: point clouds restored by stereo method.

4   3D ESTIMATION OF SOCCER SCENE

4.1   Restoration of 3D information

We estimated the 3D coordinates of the corresponding points among multiple images and the parameters of the shooting cameras by Structure from Motion (SfM) [23] and got sparse 3D-point clouds, which are the base of the 3D model. Since the spatial resolution is too sparse to generate a depth map of a dolly-in image, we applied a stereo method from multiple viewpoints based on the Patch-Match algorithm [24] to the sparse point clouds. figure 5 shows the results of applying the processes to multi-view images. We identified a depth map at an arbitrary viewpoint by increasing the distribution density of the 3D-point clouds.

4.2   Coordinate transformation

Initially, the coordinate system (e.g., the origin, the X-, Y-, and Z-axes, and the scale) of the estimated 3D-point clouds was defined by SfM. As shown in figure 6 (Left), 3D rigid transformation matrix D for conversion from the SfM coordinate system to the world coordinate system is calculated by Eq. (1). Here P_w is the 3D coordinate of an arbitrary point in the world coordinate system, and P_s represents the 3D coordinates of the same point in the SfM coordinate system:

$$P_w = DP_s, \quad D = \begin{bmatrix} sR & t \\ 0 & 1 \end{bmatrix}. \quad (1)$$

The origin of the world coordinate system is chosen as the point where two straight lines (edges) vertically intersect in the soccer field, and an object exists in this area whose size is known.

Vector t is given as the parallel translation amount from point S_o to origin O_{sfm} in the SfM coordinate system corresponding to the origin of the world coordinate system. In addition, scale s is obtained from the ratio with the size in the SfM coordinate system using an object whose size is known in the world coordinate system.

An orthonormal vector of the SfM coordinate system represented by Eq. (2) is calculated by points S_x, S_y, and S_z in the SfM coordinate system corresponding to points on the x, y, and z-axes of the world coordinate system. Rotation matrix R is obtained from the components of each vector e_i:

$$e_i = \frac{S_i - S_o}{|S_i - S_o|} \ (i = x, y, z). \quad (2)$$

(a)   (b)   (c)   (d)

Figure 7: Result of processing by proposed method:

(a) depth map projecting 3D-point clouds; (b) expansion process result;(c) mapping to depth map result; (d) image reconstruction result

5   PSEUDO-DOLLY-IN IMAGE GENERATION

5.1   Dolly-in operation of virtual camera

In the estimated 3D model, we placed a virtual camera to capture a dolly-in image. As shown in figure 6 (Right), in the camera coordinate system, the optical axis direction of the virtual camera is the same as the z-axis, the vertical direction is the y-axis, and the x-axis is orthogonal to both axes (in a right hand system). The initial position and posture of the virtual camera are set to coincide with the multi-view camera as the starting point of the dolly-in operation.

When the virtual camera is moved in toward the optical axis by the dolly-in operation, the position and orientation of the virtual camera are obtained by Eq. (3):

$$[R_d|t_d] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & c \end{pmatrix} \begin{bmatrix} R_m & t_m \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{3}$$

Here, c is a parameter that indicates the camera movement in the direction of the optical axis. $[R_m|t_m]$ expresses the position and orientation of the multi-view camera as the starting point of the dolly-in operation. $[R_d|t_d]$ expresses the position and orientation of the dolly-in camera. The focal length of the virtual camera is also set to the same value as the multi-view camera (the dolly-in operation's starting point).
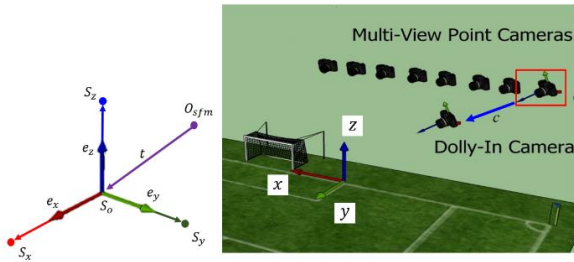

Figure 6: Left: relationship of coordinate transformation; Right: dolly-in of multi-view camera.

5.2   Generating depth map of dolly-in image from 3D model

The depth map of the dolly-in-image is generated using the 3D model created in Section 4.1. All 3D-point clouds are projected on a multi-view image as the initial step of the dolly-in operation. When more than two 3D points are projected to the same position (2D coordinate) in the dolly-in image, we removed the 3D points except the nearest one to generate a depth map of the dolly-in image. However, when the spatial distribution density of the 3D-point clouds group is insufficient, a sparse depth map is generated (figure 7(a)). In such a case, dilation processing is applied to generate a depth map without gaps (figure 7(b)).

5.3   Texture mapping onto the dolly-in image

The pixel value of the multi-view image is mapped to pixel $I_D(x, y)$ on the dolly-in image using the depth map generated from the method described in Section 5.2. By referring to the dolly-in image and the depth map value, the 3D position of the camera coordinate system that corresponds to pixel $I_D(x, y)$ is calculated by Eq. (4):

$$\begin{aligned} X &= (x - c_x) \times Z / f \\ Y &= (y - c_y) \times Z / f \\ Z &= depth. \end{aligned} \tag{4}$$

Here, c_x and c_y represent the center of the dolly-in image, and f represents the focal length of the dolly-in camera.

The position of the 3D-point clouds in the camera coordinate system is converted to the 3D position in the world coordinate system by the coordinate transformation expressed by Eq. (5):

$$p_w = A[R_d|t_d]p_c. \tag{5}$$

Here, A represents the internal parameters of the dolly-in camera, p_w represents the world coordinate system of an arbitrary 3D point, and p_c represents the 3D coordinates in the camera coordinate system. After projecting the 3D-point cloud to the multi-view image using the position information of the world coordinate system, the pixel value of the projected multi-view image is set as the value of the pixel of interest $I_D(x, y)$ of the dolly-in image. figure 7(c) shows the texture mapping result. The white region shows where the pixel values are not mapped by occlusion.

5.4   Image reconstruction processing

In the mapping process, the appearance of a region where no pixel values were obtained is compensated by image reconstruction processing. In this method, we adopted the image reconstruction method proposed by Huang et al. [25]. In this method, a pattern like the texture around the defective region is searched from the image to reconstruct the missing region from the texture information. Moreover, perspective projection distortion is also taken into consideration by using the vanishing point. Therefore, it is effective for restoring the missing field line. The result of applying this method to figure 7(c) is shown in figure 7(d). We restored the missing field line and generated higher quality images.

6 EXPERIMENT AND RESULTS

To verify the practicality of our proposed method, we captured the image of a soccer field in a large-scale, outdoor space and next describe our experiments in which we applied the proposed method to the captured images.
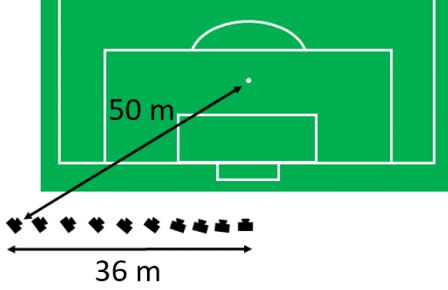

Figure 8: Camera arrangement in a large-scale space: soccer stadium

6.1 Camera settings for capturing images and initial environment for applying proposed method

We captured the images of a youth soccer match at a professional soccer stadium. In the environment for capturing the images, we arranged ten cameras at the seats behind the soccer stadium's goal at intervals of about 4 m (figure 8). The region that captures the images is near the goal. Multi-videos were captured using digital single-lens reflex cameras (Canon EOS 5D Mark-II,) with 5,616 x 3,744 pixel resolution at a focal of length 50 mm.

Corresponding point information was calculated from the captured multi-view images to generate camera calibration and 3D-point clouds. Then we set the world coordinate system of the captured scene. In the goal region, many 3D-point clouds are generated on the field line. Since the field line size is defined by official soccer regulations, the conditions described in Section 4.2 are satisfied. Therefore, as shown in figure 6 (lower), we set the origin of the world coordinate system and each axis. The soccer goal direction from the origin is the x-axis, the center direction of the soccer field is the y-axis, and its vertical upward direction is the z-axis.

The upper part of figure 9 shows a multi-view image that captured the image. The 3D estimation result using the captured multi-view image is shown in the lower part of figure 9. In this demonstration experiment, we confirmed the proposed method's applicability using the 3D model of the lower part of figure 9. In comparison experiments, two image-sets were used: a digital zoom image-set and a dolly-in image-set generated from the IBMR image-set. We confirmed the effectiveness of the proposed method using these two image-sets.

6.2 Pseudo-dolly-in production experiment

The upper part of figure 9 shows the sequence of the multi-view images. Demonstration experiments were performed on one frame in the multi-view video. The targeted frame is the image enclosed by the black dotted line in the upper left of figure 9. For this image, we generated digital zoom video, IBMR video, and pseudo-dolly-in video by the proposed method (figure 10). The upper row of figure 10 shows the digital zoom video, the middle figure shows the IBMR video, and the lower figure shows the pseudo-dolly-in video by the proposed method. In this sequence, when the proposed method was applied, we confirmed that pseudo-dolly video is generated.

We compared the digital zoom and IBMR image-sets. Digital zoom images cannot reproduce motion parallax because no depth is understood (figure 10 ※1). In the IBMR images, motion parallax can be reproduced using depth information (figure 10 ※2).

Next, we compared the IBMR and the pseudo-dolly-in image-sets. As for the pseudo-dolly-in image by the proposed method, motion parallax was reproduced, like the IBMR image using depth information (figure 10 ※3).

Next we checked the quality of the three image-sets (digital zoom image, IBMR image, and pseudo-dolly-in). An enlarged view of each method is shown at the bottom of figure 10. The image quality of ※1 and ※3 is equally high. On the other hand, since the image quality of ※2 depends on the 3D estimation accuracy, the quality is relatively low.

Table 1 summarizes the experimental results of image quality and the sense of immersion in each method. The image quality of ※ 3 is high, like ※ 1. Moreover, since motion parallax is reproduced based on 3D information like ※ 2, ※ 3 successfully obtained a sense of immersion that resembles moving in space. Furthermore, the image quality is maintained at more than ※2 because the proposed method maps the pixel values of the original multi-view image. Then we applied reconstruction processing to the missing region.

The above results confirm that motion parallax was reproduced in the image generated by the proposed method. We presented a sense of immersion that surpassed digital zoom.

Table 1: Experimental results on image quality and sense of immersive

	Image quality	Sense of immersion
Digital zoom image	◯	✕
IBMR image	✕	◯
Pseudo-dolly-in image (our proposed method)	◯	◯

7 CONCLUSION

We proposed a method to generate pseudo-dolly-in video using image reconstruction processing for multi-view images. We reconstructed a 3D model from multi-view images and used the depth information obtained from the restored 3D model. Therefore, images were generated that reproduced high-quality motion parallax. In addition, reconstruction processing was applied, and an image was generated that maintained visual continuity as much as possible and applied to multi-view capturing images in a large-scale space. We demonstrated the proposed method's effectiveness. As a result of comparative experiments, we concluded that the image quality of the digital zoom video and the pseudo-dolly-in video by the proposed method had equally high quality. The proposed method's higher image quality exceeds the dolly-in image generated from Image-Based Modeling and Rendering methods.

Figure 9: Sequence used for comparison experiment:

Upper: multi-view image capturing the image. Lower: result of reconstruction of 3D information using captured multi-view image.

Digital zoom   ※1

Image-Based Modeling
and Rendering   ※2

Proposed method
Pseudo-dolly-in video   ※3
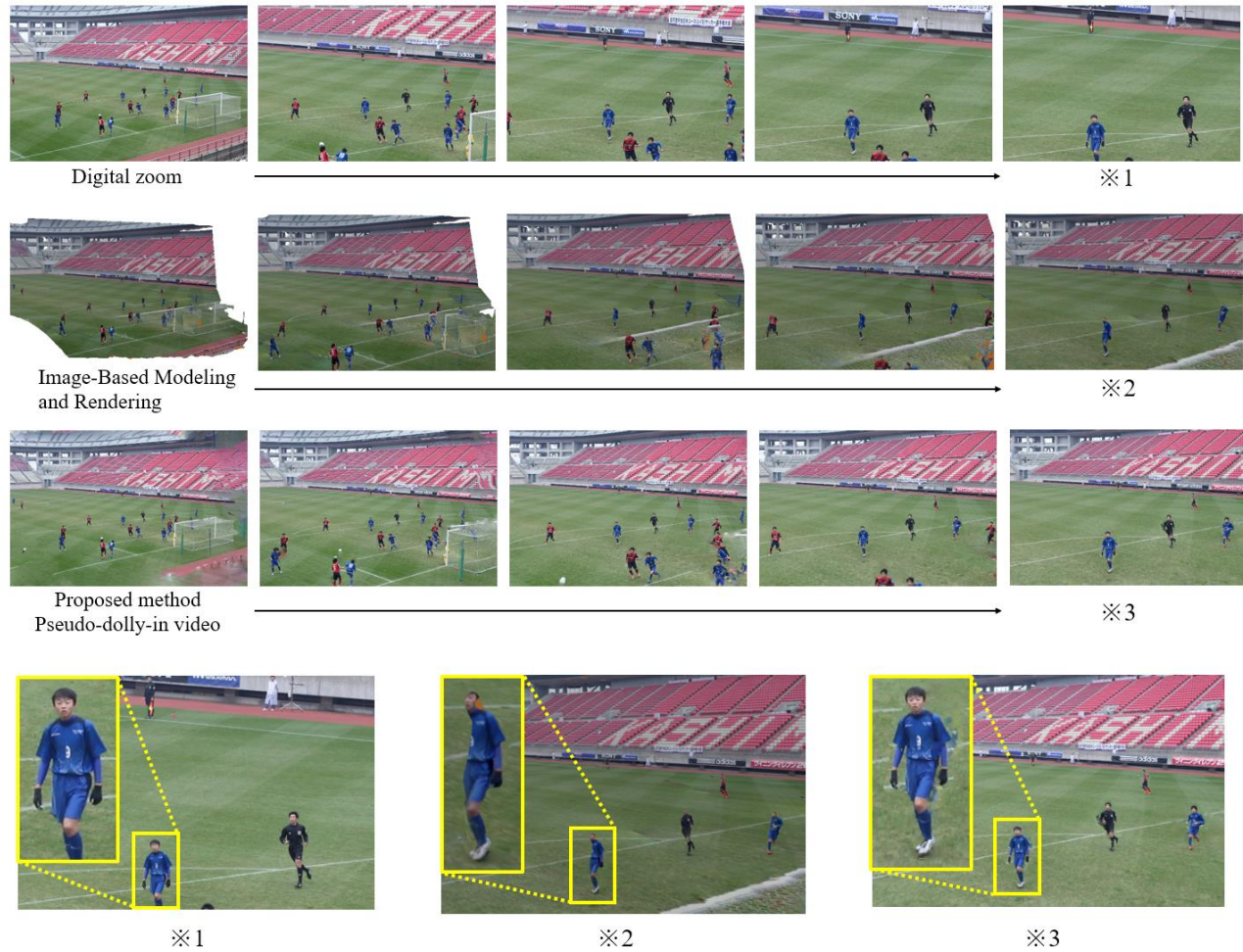
※1   ※2   ※3

Figure 10:   Result images for soccer scene. Targeted frame is image enclosed by black dotted line in upper left of figure 9.

Upper: digital zoom video. Middle: IBMR video. Lower: pseudo-dolly-in video by proposed method. Bottom: enlarged view of each method. ※ corresponds to each method.

332

REFERENCES

[1] T. Kanade, P. Rander, and P. J. Narayanan. Virtualized Reality: Constructing Virtual Worlds from Real Scenes. IEEE MultiMedia, Vol. 4, No. 1, pp. 34-37, 1997.

[2] T. Koyama, I. Kitahara, and Y. Ohta. Live Mixed-Reality 3D Video in Soccer Stadium. Proceedings of IEEE 2003 International Symposium on Mixed and Augmented Reality (ISMAR), pp. 178-187, 2003.

[3] Liu, Yebin, Qionghai Dai, and Wenli 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] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen: On benchmarking camera calibration and multi-view stereo for high resolution imagery. 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2008.

[5] Y. Furukawa and J. Ponce: Accurate, Dense, and Robust Multiview Stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 8, pp. 1362-1376, 2010.

[6] Simon Fuhrmann, Michael Goesele: Floating scale surface reconstruction. ACM Transactions on Graphics (TOG), Vol. 33, No. 46, 2014.

[7] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz: Multi-View Stereo for Community Photo Collections. IEEE 11th International Conference on Computer Vision (ICCV), pp. 1-8, 2007.

[8] A. Hornung and L. Kobbelt: Hierarchical Volumetric Multi-view Stereo Reconstruction of Manifold Surfaces based on Dual Graph Embedding. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 503-510, 2006.

[9] M. Jancosek, A. Shekhovtsov, and T. Pajdla: Scalable multi-view stereo. IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 1526-1533, 2009.

[10] A. Locher, M. Perdoch, and L. V. Gool: Progressive Prioritized Multi-view Stereo. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3244-3252, 2016.

[11] Shin, Tetsuya, et al. A comparison between two 3d free-viewpoint generation methods: Player-billboard and 3d reconstruction. 2010 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video, IEEE, 2010.

[12] Nao Akechi, Itaru Kitahara, Ryuuki Sakamoto, and Yuichi Ohta: Multi-Resolution Bullet-Time Effect. ACMSIGGRAPH Asia 2014, 2014.

[13] Ding Chen and Ryuuki Sakamoto. Optimizing Infinite Homography for Bullet-Time Effect. ACMSIGGRAPH Asia 2014.

[14] Tomoyuki Mishina andYuichi Iwadate: Bullet Time Using Multi-Viewpoint Robotic Camera System, 11th European Conference on Visual Media Production (CVMP), Article No. 1, 2014.

[15] Nakanishi, Hideyuki, Yuki Murakami, and Kei Kato. Movable cameras enhance social telepresence in media spaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2009.

[16] Setlur, Vidya, et al. Automatic image retargeting. Proceedings of the 4th international conference on Mobile and ubiquitous multimedia. ACM, 2005.

[17] Belton, John. The bionic eye: Zoom esthetics. Cineaste vol. 11, No. 1, pp. 20-27, 1980.

[18] Hideyuki Nakanishi, Yuki Murakami, Kei Kato, "Movable cameras enhance social telepresence in media spaces," SIGCHI Conference on Human Factors in Computing Systems, pp. 433-442, 2009.

[19] A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, pp. 150-162, 1994.

[20] T. Kanade, et al. Eye Vision. http://www.ri.cmu.edu/events/sb35/tksuperbowl.html, 2001.

[21] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. ACM Transactions on graphics (TOG). Vol. 26. No. 3. ACM, 2007.

[22] Shen, Jianbing, Dapeng Wang, and Xuelong Li. Depth-aware image seam carving. IEEE transactions on cybernetics Vol. 43, No. 5, pp. 1453-1461, 2013.

[23] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. ACM SIGGRAPH Transactions on Graphics, vol. 25, no. 3, pp. 835-846, 2006.

[24] Barnes, Connelly, et al. PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics-TOG Vol. 28, No. 3, 2009.

[25] Huang, Jia-Bin, et al. Image completion using planar structure guidance. ACM Transactions on Graphics (TOG), Vol. 33, No. 4, 2014.