Pan-zoom Motion Capture in Wide Scenes using Panoramic Background

Masanobu Yamamoto
Department of Information Engineering, Niigata University, Niigata, Japan

Keywords: Multi-view Videos, Panoramic Background Image, Pan and Zoom, 3D Motion, Tracking.

Abstract: Measuring a subject three-dimensionally from multiple cameras, the measurable area is a common field of view from cameras. When the subject goes out of the field of view, the cameras must follow the subject. In this research, the viewpoint of the camera keeps to be fixed and the pan and zoom functions are used to operate the cameras so that the subject body could be always shot near the center of the image. The problem is camera calibration. Our approach is to use a panoramic image. Each camera pans in advance to take a background image and create a panoramic image of the background. Then, the background image around the subject body is collated with the panoramic image, the pan rotation angle and the zoom ratio are obtained from the matching position, and the camera is calibrated. The body motion is captured from the multi-view motion image using the camera parameters obtained in this way. Since the viewpoint of the camera is fixed, the shooting range is not so wide, but it is still possible to capture an athlete's floor exercise in the gymnasium.

1 INTRODUCTION

Motion capture from video image does not constrain the subject's body and natural movement can be measured without giving the subject the consciousness of being measured. It is possible to measure the 3D movement of the body from an even single camera view, but it is a guess rather than a measurement because the depth information is missing. By using the image from the multi-view cameras, it is possible to accurately measure the 3D movement of the body.

When measuring the 3D movement by multiple cameras, the measurable area is only the common field of view from cameras. For example, at the Fig.1 (a), the common field of view from the four cameras is shown by the gray area. To measure even if a person goes out of this area, one can pan the camera as shown in the Fig.1 (b). However, if the body gets too close to the camera, a part of the body will be out of sight. Also, as the body goes away from the camera, the apparent size becomes smaller and the resolution of the body image becomes lower. Therefore, by adding a zoom to the pan, the imaging range can be further expanded as shown at the Fig.1 (c).

The problem here is a camera calibration. If the camera is fixed, it is sufficient to present the calibration object only once within the common field of view. Otherwise, it is not possible to place the calibration object during pan and zoom. Our idea is to use a panoramic image of the background. Therefore, each camera is pan-rotated in advance to take a background image and create a panoramic image of the background. When capturing motion, let's pan and zoom the cameras so that the body could be always shot near the center of the image. Then, the background image around the body is collated with the panoramic image, and the pan rotation angle and the zoom ratio can be obtained from the matching position. The movement of the body is captured from the multi-view videos using the camera parameters obtained in this way.

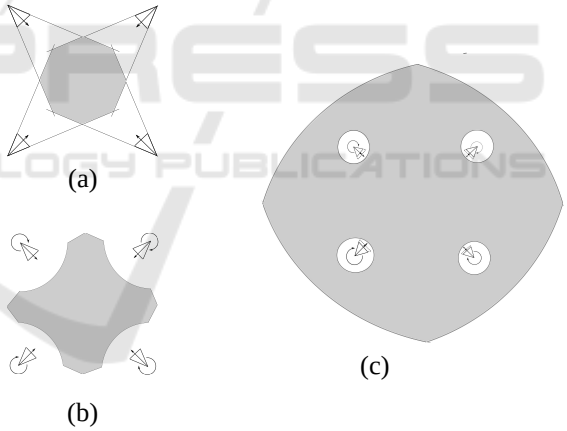

Figure 1: Common field of view (gray area). (a) Fixed cameras, (b) Pan cameras, (c) Pan-zoom cameras.

Yamamoto, M.
Pan-zoom Motion Capture in Wide Scenes using Panoramic Background.
DOI: 10.5220/0010815600003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 1: GRAPP*, pages 163-171
ISBN: 978-989-758-555-5; ISSN: 2184-4321

2 RELATED WORKS

Motion capture from multi-view cameras began in the late 1990s. (Sundaresan and Chellappa, 2005) reviews on motion capture from multi-view video images up to around 2005.

While multi-view cameras can measure accurate 3D positions, they can only measure in common field of view between cameras. Increasing the number of cameras is one solution (Joo et al., 2015), but it needs a high cost of entire facility. Therefore, the camera had better be made to follow the movement of the body. (Rhodin et al., 2016) attached a stereo camera to the head and measured the movement of the body below the neck. Since the camera moves with a person, the measurement range is not limited to the movement of the person. However, although the relative movement of the body part with respect to the root of the body can be recovered, the movement of the root is the movement with respect to the camera coordinate system, not the movement with respect to the world coordinate system. The camera in motion should be calibrated in the world coordinate system.

If the camera moves a lot, it is effective to use external sensors such as IMU and/or GPS (Xu et al., 2016; Nageli et al., 2018; Saini et al., 2019). The use of external sensors is effective (Kurihara et al., 2002; Ukita and Matsuyama, 2005) even when the viewpoint of the camera is fixed as in our research. In such case, a camera calibration is possible just from the background image without any external sensors. In fact, many PTZ camera self-calibration techniques (Shum and Szeliski, 2000; Sinha and Pollefeys, 2006; Wu and Radke, 2013) have been proposed to create accurate background panoramic images. However, in the motion capture, the self-calibration techniques cannot be used because the background image contains a body image. Therefore, we decide to create a panoramic image of the background in advance and calibrate the camera by matching the panoramic image with the image including the body. The idea of using a panoramic image for calibration can be found in (Cannelle et al., 2010), but it's just a simulation.

This paper treats two types of image matching. One is when detecting the overlap in the image sequence obtained by pan rotation to make a panoramic image. Another is when the background of the body image is collated with the panoramic image. In the early days, matching was performed based on the image intensity (Shum and Szeliski, 2000), but since SIFT (Lowe, 2004), feature-based matching has become popular. Initially, we used SURF (Bay et al., 2008), but switched to KAZE (Alcantarilla et al., 2012), which can maintain the boundary of the area more clearly. Furthermore, AKAZE (Alcantarilla et al., 2013) is currently used to improve the calculation speed.

3 PANORAMIC BACKGROUND

Let's show a cylindrical panorama, which is an ease of construction (Shum and Szeliski, 2000) and is sufficient for our motion capture use.

3.1 Camera Coordinate System

We show a perspective camera model and a cylindrical coordinate system that projects all directions.

3.1.1 A Perspective Camera Model

Let $O : X,Y,Z$ be an orthogonal coordinate system (see Fig.2), in which O is the projection center, the Z axis is an optical axis, and the projection plane is placed at the position f on the optical axis. The plane is perpendicular to the optical axis. Let the intersection of the plane and the optical axis be the origin of the projection plane coordinates $o : x,y$. The 3D position $(X,Y,Z)^T$ in space is projected onto the 2D position $(x,y)^T$ of the projection plane as follows,

$$\begin{cases} x = f\dfrac{X}{Z} \\ y = f\dfrac{Y}{Z} \end{cases} \tag{1}$$

Let the image coordinates on the projection plane be (u,v), the origin of the projection plane be (u_0,v_0), and the 2D scale transformations from the projection plane coordinate axes to the image plane coordinate axes be k_u and k_v, respectively. The relationship between both coordinates is

$$\begin{cases} u = k_u x + u_0 \\ v = k_v y + v_0 \end{cases} \tag{2}$$

3.1.2 Cylindrical Camera Model

The perspective camera has a limited field of view, while a cylindrical camera can shoot omnidirectional scenes. The cylindrical camera assumes to share the projection center with the perspective camera, and sets a cylinder whose axis is a straight line passing through the projection center and is perpendicular to the optical axis as the projection surface. The point at which the line of sight from this projection center to the point in the scene intersects the cylindrical surface is defined as the projection point. Unfolding

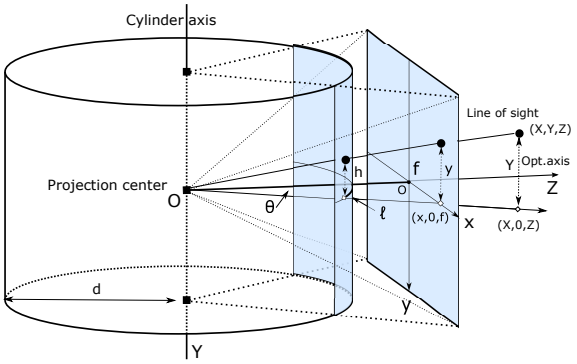

Figure 2: Perspective and cylindrical projections. Assuming that both share a viewpoint.

this cylindrical surface, all points in the scene are projected on the 2D image.

Since the perspective camera cuts out a part of the entire circumference, it is necessary to concatenate many perspective projecting images in order to obtain a cylindrical image of the entire circumference. At this time, each perspective projection image is transformed into a cylindrical image and pasted together on the cylinder

Let the radius of the cylinder be d (see Fig.2), and the length of the arc in the circumferential direction starting from the intersection of the YZ plane and the cylinder be l. Let the height in the direction of the cylinder axis starting from the intersection of the XZ plane and the cylinder be h, and (l,h) be the coordinates of the cylinder surface.

Let consider the projection of the line of sight seeing the 3D point $(X,Y,Z)^T$ onto the XZ plane. If the angle between the projection of the line of sight and the optical axis is θ, the arc length is $l = d\theta$. Also, since $\theta = \tan^{-1} x/f$, from the eq.(2), we have

$$l = d\tan^{-1}\frac{u-u_0}{k_u f} \tag{3}$$

On the other hand, considering a sector constructed from the line of sight and its projection,

$$h = d\frac{y}{\sqrt{x^2+f^2}}$$

Substituting the eq.(2) onto it, we have

$$h = d\frac{k_u}{k_v}\frac{v-v_0}{\sqrt{(u-u_0)^2+k_u^2 f^2}} \tag{4}$$

Furthermore, let's transform the cylindrical surface (l,h) into an unfolded image plane (r,s). If the origin of the cylindrical surface is (r_0,s_0) on the image plane and the transformation scale of each coordinate axis is k_r and k_s, the transformation is

$$\begin{cases} r = k_r d\tan^{-1}\dfrac{u-u_0}{k_u f} + r_0 \\ s = k_s d\dfrac{k_u}{k_v}\dfrac{v-v_0}{\sqrt{(u-u_0)^2+k_u^2 f^2}} + s_0 \end{cases} \tag{5}$$

Let $d = 1$, $k_r = k_u$, $k_s = k_v$, we have

$$\begin{cases} r = k_u\tan^{-1}\dfrac{u-u_0}{k_u f} + r_0 \\ s = k_u\dfrac{v-v_0}{\sqrt{(u-u_0)^2+k_u^2 f^2}} + s_0 \end{cases} \tag{6}$$

where f, k_u, k_v, u_0 and v_0 are internal parameters of the perspective camera and can be obtained by camera calibration. The r_0 and s_0 are the paste positions on the unfolded image plane. The calibration method for camera in motion will be described later.

3.2 Shooting Environment

Video images are shot with four video cameras (CANON, XH-G1) at the four corners on the floor. Fig.3 left shows the camera layout. An external synchronization signal is supplied to each camera by wire. In addition, one camera is named as a master camera, and the SMPTE time code (TC) of the master camera is supplied to the remaining cameras by wire.

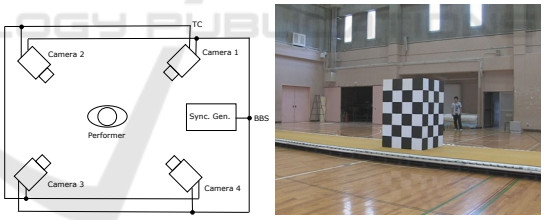

Figure 3: Left: Arrangement of synchronized multi-view cameras, Right: Camera calibration object.

Videos are recorded on DV tapes. We transfer the videos to the HDD while reading the TC on the DV tape with a video editor (Grass Valley co., REXCEED model 3100). Time series of a set of four multi-viewpoint images taken at the same time can be obtained.

3.3 Elimination of Parallax

So as to keep the moving subject in the center of the field of view, we pan-rotate the camera. The camera is fixed to a tripod head, and rotate the pan head to change the direction of the camera. At this time, the projection center of the camera is not necessarily on the axis of rotation.

Sandwiching a slide head between the camera and the pan head, the projection center aligns with the rotation axis by sliding the camera back or forth. When they align, the foreground always appears at the same position in the background, regardless of camera rotation (Cannelle et al., 2010).

3.4 Calibration at Reference Pose

Before panning the camera, we place the camera calibration object shown on the right of the Fig.3 instead of the performer in the common field of view of all the cameras. The pose of the camera at this time is called the reference pose. Let the coordinate system attached to this calibration object be the world coordinate system, and obtain the parameters of each camera with respect to the world coordinate system. Of these, the internal parameters f, k_u, k_v, u_0 and v_0 are used to convert the perspective projection image into a cylindrical image. The camera calibration used Tsai's method (Tsai, 1986).

3.5 Making a Panoramic Image

After removing the calibration object, slowly pan-rotate the camera to shoot a background movie, and connect the image frames to make a panoramic image.

The image taken by perspective camera is transformed into a cylindrical image using the eq.(6). Procedure is : Pick up one frame from the background movie and paste it on the cylinder. Compare the second frame with the first frame to find the displacement between the two frames by AKAZE (Alcantarilla et al., 2013). Move the second frame using this displacement from the first frame on the cylindrical surface, and paste it where the pixel value is not yet written. Repeat this procedure for each new frame.

The panoramic images unfolded from the cylindrical surface are shown in the Fig.4. When the camera at the reference pose was calibrated in the subsection 3.4, the vertical and horizontal angles of view of the camera are obtained, and the rotation angle per one pixel is also derived. The horizontal and vertical axes of the panoramic image can be described as the pan and tilt rotation angles, respectively.

4 CALIBRATION IN PAN-ZOOM

The subsection 3.4 described camera calibration at the reference position. This section will describe the calibration of the camera in pan-zoom.

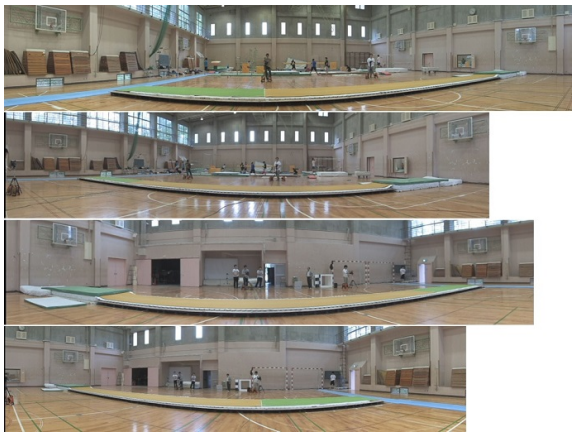

Figure 4: Panorama images from cameras 1, 2, 3 and 4 in order from the top.

4.1 Matching with Panoramic Images

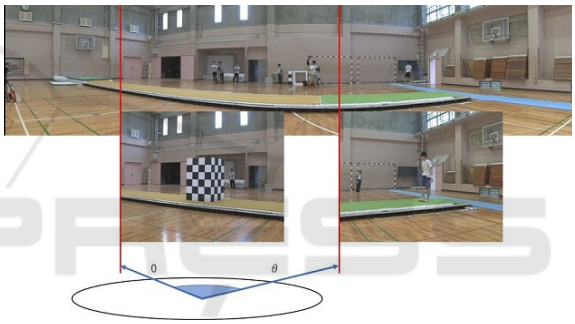

Figure 5: Two positions on the panoramic image from camera 1, that match with the camera calibration image taken at the reference pose and the image taken after panning, respectively.

The camera calibration image taken in the reference position is collated with the panoramic background image. The matched position becomes an origin on the panoramic image. Fig.5 shows the panoramic image of camera 4 selected from the Fig.4, on which the origin is denoted by matching with the calibration image. We also obtain the matched position of the image shot by panning and zooming camera with the panoramic image as shown in Fig.5.

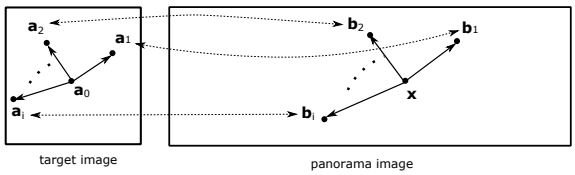

Figure 6: The left figure denotes a cylindrical image transformed from the perspective image. The right figure denotes the panoramic image of the background.

When the feature point correspondences between the target image and the panoramic image and the origin of the target image are given, the problem is to obtain the corresponding destination of the origin on the panoramic image, and zoom ratio. Fig.6 shows corresponding pair, $\mathbf{a}_i = (r_i, s_i)^T$ and $\mathbf{b}_i = (p_i, q_i)^T$, the origin of the target image, $\mathbf{a}_0 = (r_0, s_0)^T$, and the corresponding destination, $\mathbf{x} = (p_0, q_0)^T$.

4.2 Pan /Zoom Camera Calibration

According to eq.(6), the transformation from a perspective image to a cylindrical image is

$$\begin{cases} r = k_u \tan^{-1} \dfrac{u-u_0}{k_u f} + r_0 \\ s = k_u \dfrac{v-v_0}{\sqrt{(u-u_0)^2 + k_u^2 f^2}} + s_0 \end{cases} \tag{6}$$

When the focal length f is changed to η times by zooming, the relative position $(u-u_0, v-v_0)^T$ of the projection point $(u,v)^T$ for the image center point $(u_0,v_0)^T$ also changes to $\eta(u_f-u_0, v_f-v_0)^T$ as shown in the Fig.7. Here, $(u_f,v_f)^T$ is the projection point before zooming. Even if f and $(u-u_0, v-v_0)^T$ in the eq.(6) are replaced with to ηf and $\eta(u_f-u_0, v_f-v_0)^T$, the projection position on the cylindrical surface does not change since the zoom ratio η is canceled by the numerator and denominator. This is natural because the focal length of the perspective transformation has nothing to do with the cylindrical transformation.

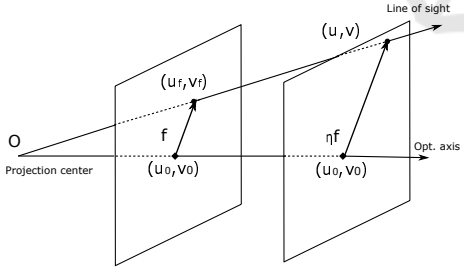

Figure 7: Change of the relative position on the perspective image after zoom up.

However, when a perspective image taken by zooming and panning is transformed into a cylindrical image by the eq.(6), the effect of zooming appears on the cylindrical image. This is because the focal length is not multiplied by the zoom ratio η. On the contrary, the zoom ratio can be obtained from this effect.

The transformation to cylindrical coordinates before zooming is given by the eq.(6), which is represented by the feature points on the panoramic image.

$$\begin{cases} p - p_0 = k_u \tan^{-1} \dfrac{u_f - u_0}{k_u f} \\ q - q_0 = k_u \dfrac{v_f - v_0}{\sqrt{(u_f-u_0)^2 + k_u^2 f^2}} \end{cases} \tag{7}$$

On the other hand, since the points affected by the ratio η by zooming are represented by the feature points on the cylindrical image, we have

$$\begin{cases} r - r_0 = k_u \tan^{-1} \dfrac{u-u_0}{k_u f} \\ s - s_0 = k_u \dfrac{v-v_0}{\sqrt{(u-u_0)^2 + k_u^2 f^2}} \end{cases} \tag{8}$$

Assuming that the image center (u_0,v_0) and the scale k_u does not change before and after zooming, we have the projection position as seeing the Fig.7.

$$\begin{cases} u - u_0 = \eta(u_f - u_0) \\ v - v_0 = \eta(v_f - v_0) \end{cases} \tag{9}$$

The position after zooming can be observed, but the position before zooming cannot be observed. Setting ν with $1/\eta$,

$$\begin{cases} u_f - u_0 = \nu(u - u_0) \\ v_f - v_0 = \nu(v - v_0) \end{cases} \tag{10}$$

Substitute this relationship into the eq.(7) and compare the ratios of the left and right sides of the eqs. (7) and (8), we have

$$\begin{cases} \dfrac{p-p_0}{r-r_0} = \dfrac{\tan^{-1} \dfrac{\nu(u-u_0)}{k_u f}}{\tan^{-1} \dfrac{u-u_0}{k_u f}} \\ \dfrac{q-q_0}{s-s_0} = \nu \sqrt{\dfrac{\nu^2(u-u_0)^2 + k_u^2 f^2}{(u-u_0)^2 + k_u^2 f^2}} \end{cases} \tag{11}$$

This equation is a non-linear equation with (p_0,q_0) and ν as unknowns.

Let's linearize this nonlinear equation. Using the approximate equation $\tan^{-1}\theta \cong \theta$ (suppose $|\theta|$ to be small), the first equation of the eq.(11) is

$$\frac{p-p_0}{r-r_0} = \nu \tag{12}$$

Since ν is near 1, let's set $\nu = 1+\delta$, and expand the right side $\sqrt{*}$ of the second equation of eq.(11) with δ into the Taylor series. Ignoring the second and higher order small terms, we have

$$\sqrt{\frac{(1+\delta)^2(u-u_0)^2 + k_u^2 f^2}{(u-u_0)^2 + k_u^2 f^2}} \cong 1 + \frac{(u-u_0)^2}{(u-u_0)^2 + k_u^2 f^2}\delta \tag{13}$$

Furthermore, supposing $(u-u_0)^2 \ll k_u^2 f^2$, since the second term on the right side is a higher-order small term, we can ignore it. we have

$$\frac{q-q_0}{s-s_0} = \nu \tag{14}$$

These approximations hold near the center of the image, where the cylindrical image almost equals the perspective image.

Arranging the eqs.(12) and (14) as a system of linear equations,

$$\begin{pmatrix} r-r_0 & 1 & 0 \\ s-s_0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu \\ p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix} \tag{15}$$

By solving the above equations constructed from multiple corresponding points, the zoom ratio $\eta = 1/\nu$ and the position $(p_0, q_0)^T$ on the panoramic image can be obtained. With this solution as the initial solution, an exact solution can be obtained by solving the nonlinear eqs.(11), but in this paper, the initial solution is used as the final solution. Of the matched position $(p_0, q_0)^T$, q_0 is affected by tilt rotation while p_0 is affected by pan rotation. Tilt pose can also be calibrated, but in our experiment, the tilt rotation was not performed due to the limit of manual operation of the camera.

4.3 Experiments of Pan-zoom Camera Calibration

Let's show the calibration of pan-zoom cameras in the experimental laboratory scene. At an initial step, a calibration object is shot in the center of image, and check where the calibration object positions. The reference pose of the camera is obtained from the image of this calibration object. When the model of the calibrated object is projected using the obtained camera parameters, one can see that the blue line wireframe model matches the object, as shown on the left in Fig.8.

Figure 8: Left: Calibration result at reference pose, Right: Calibration result after pan / zoom.

A panoramic image of the background is created from the movie obtained by pan-rotating the camera after removing out the calibration object from the scene. Next, return the calibration object to its original position, then pan the camera to the right to zoom in and shoot the calibration object. The shot target image is shown in the right of Fig.8.

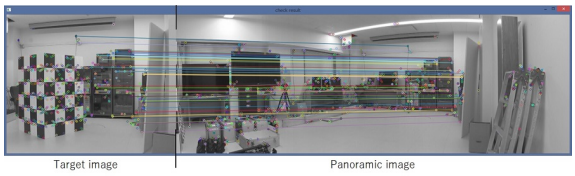

Figure 9: The left end is the image of the target taken after pan-zooming the camera. The right part is a panoramic image of the background. The result of matching both by AKAZE (Alcantarilla et al., 2013) is shown by straight lines connecting the corresponding feature points.

The target image is collated with the panoramic image. The matching result is shown in Fig.9. AKAZE (Alcantarilla et al., 2013) extracts feature points which are indicated by small circles. Of these feature points, matched pairs are associated with line segments. We can obtain the focal length f from the zoom ratio η calculated from the associated pairs, and the pan rotation angle from the position (p_0, q_0) on the panoramic image, and update the camera parameters.

Strictly speaking, when zooming, the lens system moves back and forth, so the projection center may move and a parallax appears. Moreover, we do not consider a distortion of the camera lens. We were worried that these effects would reduce the accuracy of camera calibration. Using the updated camera parameters, overlay a model of the calibrated object on the target image. Fig.8 right denotes the blue line wireframe model drawn on the calibration object, indicating that the camera is calibrated almost correctly.

5 TRACKING

After obtaining the camera parameters, the next issue is to estimate the body pose in motion. Frame by frame pose estimation means tracking the body. The tracking is performed by matching an articulated body model with each body image of the sequence. We performed the tracking based on the existing differential approach (Yamamoto et al., 2014; Kobayashi et al., 2018), where the body model has to be manually matched with the body image at several keyframes including the start and end frames to eliminate a tracking drift (Yamamoto, 2005). Tracking experiments can be seen in videos of (Kobayashi and Yamamoto, 2015; Kobayashi and Yamamoto, 2018).

6 EXPERIMENTS

Experiments show the effects of camera panning and zooming, which are the two advantages of the method proposed in this paper.

6.1 Panning Effects

Let's show panning effects by capturing motion of a gymnastic athlete in floor exercise. After shooting the exercise, motion capture is performed by offline. Captured results are overlapped on the panoramic image from camera 3 as shown in the upper row of Fig.10, where results sampled from 171 frames in total are represented by the skeleton model. The red-only skeleton represents the pose given in keyframes, and the color-coded skeleton for each part represents the pose calculated by tracking in between keyframes. The horizontal axis of the panoramic image denotes pan rotation angle. The middle row of Fig.10 denotes the corresponding tracking images in movie from camera 3. The red arrow indicates position matching each target image with the panoramic image. The camera 3 is panning from right to left. The lower row in Fig.10 shows the corresponding CG image of the athlete reconstructed from a camera view different from any four cameras.

6.2 Zooming Effects

A pedestrian is captured with pan-zoom cameras. At this time, if the pedestrian moves away, the focal length is lengthened to zoom in, and if it gets too close, the focal length is shortened to zoom out, and the entire body is always adjusted so that it fits within the image frame. This pan / zoom operation was performed manually.

The top views of the movement of the camera and the path of the pedestrian are shown in Fig.11 which shows 4 frames out of 150 tracking frames in total. The focal length is shown in the Fig.12. In this figure, the reference focal length when the camera calibration object was shot is also drawn with a dotted line. The calibration object was placed near the center of the rectangle with the camera position as the apex.

The pedestrian starts from the vicinity of the camera 3 and its path draws an arc to approach the camera 2. Initially, the camera 3 sets the focal length shorter than the reference focal length in order to capture a nearby pedestrian, but increases the focal length as the pedestrian moves away. On the other hand, in the camera 2, the focal length is shortened because the pedestrian gradually approaches. For cameras 1 and 4, the focal length is always longer than the reference focal length because the pedestrian is farther than the position where the calibration object was placed.

The Fig.13 shows the images taken with the varying focal length and the images taken with the fixed focal length for cameras 3 and 4. For each camera, the upper row is the images at the time of variable focus, and the lower row is the images at the time of fixed focus. The shooting time is the first, 50th, 100th, and 148th frames from the left. The captured skeleton overlaps on the body image. The fixed focus image in the lower row is estimated from the variable focus image on the upper row. When the variable focal length is f, the projected position $(x,y)^T$ of the position $(X,Y,Z)^T$ in space is given by the eq.(1). Supposing the fixed focal length to be f_0 which is the reference focal length, the projection position (x',y') is given by

$$\begin{cases} x' = f_0 \dfrac{X}{Z} \\ y' = f_0 \dfrac{Y}{Z} \end{cases} \tag{16}$$

From the eqs.(1) and (16), the correspondence from the variable focus image to the fixed focus image is

$$\begin{cases} x' = \dfrac{f_0}{f} x \\ y' = \dfrac{f_0}{f} y \end{cases} \tag{17}$$

The fixed focus image is made by this transformation from the variable focus image.

According to Fig.12, since the ratios f_0/f of the cameras 1 and 4 are always less than 1, the variable focus image is reduced to make a fixed focus image. Even objects that look small in the fixed focus image can be enlarged and easily measured by making the focal length variable like in camera 4 of Fig.13. On the other hand, with cameras 2 and 3, the ratio f_0/f is sometimes larger than 1, so the fixed-focus image extends beyond the angle of view of the camera. In fact, in the fixed-focus images of the 1st and 50th frames of the camera 3, the toes of the pedestrian are hidden outside of the view. By adjusting the focal length, the body image can be kept within the field of view.

7 CONCLUSION

This paper proposed a motion capture system by pan-zoom cameras. The camera calibration is performed by matching the panoramic image of the background taken in advance with the background of the body image. Using the calibrated multi-view camera system, it is possible to capture an athlete's floor exercise in the wide gymnasium.

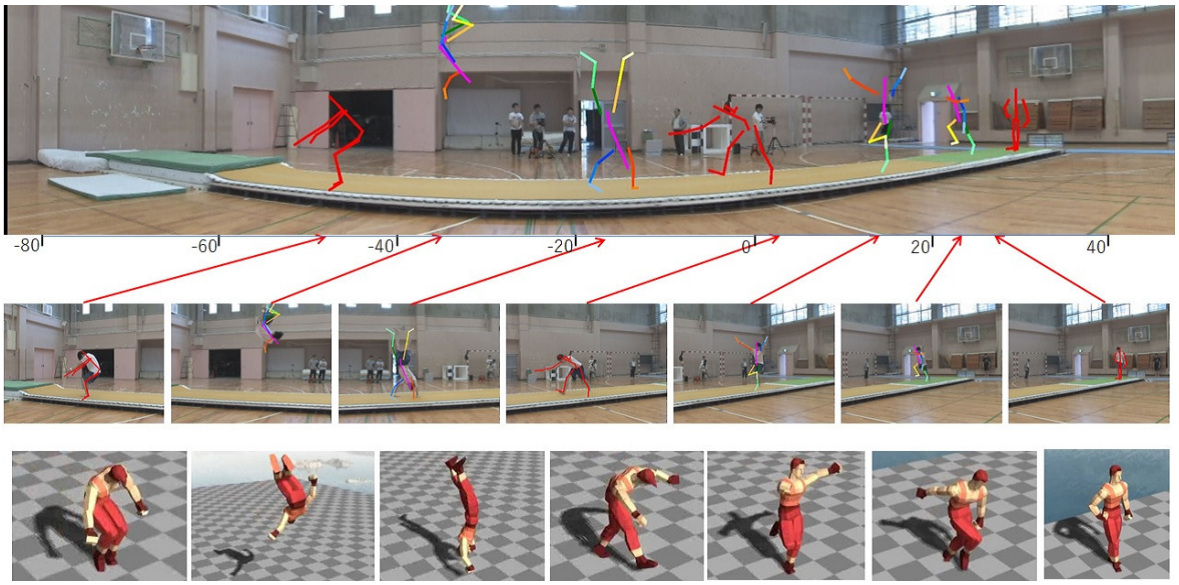

Figure 10: Upper row: The sampled results of motion capture are overlapped on the panoramic images as a stick model. Middle row: Arrange sampled tracking results which match the panoramic image at position indicated by red arrow. Lower row: Corresponding CG images of the athlete reconstructed from different camera view.

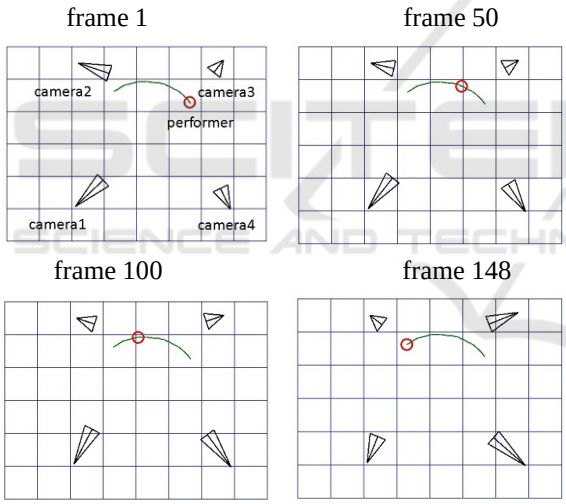

Figure 11: The path of the pedestrian and the pan and zoom of the camera in tracking on a grid with $2m$ per one unit. The camera is represented by an isosceles triangle of which apex and height denote a viewpoint and focal length, respectively. The pedestrian draws a green arc on which a red circle denotes a current position.

In the future, we plan to develop a more precise calibration method that takes into account lens distortion and a motion parallax when zooming, and compare it with other methods (Shum and Szeliski, 2000; Sinha and Pollefeys, 2006; Wu and Radke, 2013) such as self-calibration.

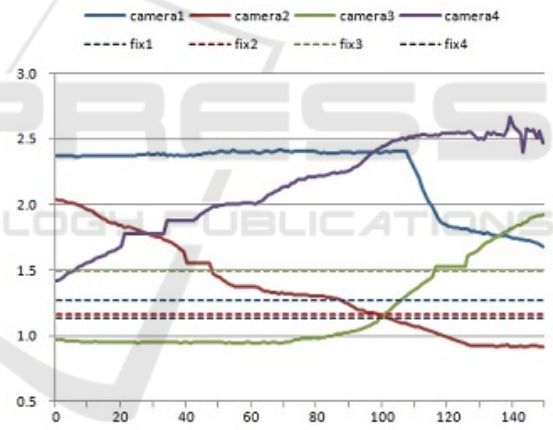

Figure 12: The focal length of each camera. A dot line denotes the focal length by camera calibration at the reference pose. The horizontal axis denotes the time by frame number. The vertical axis denotes the focal length by $\times$ 640 pixels.

ACKNOWLEDGEMENTS

I would like to thank Mr. Daisuke Kobayashi, Mr. Takashi Igarashi and Mr. Shinya Suzuki for their contributions for the early works, and also thank Mr. Shotaro Kawaguchi for his gymnastic performance. This work was supported by JSPS KAKENHI Grant Numbers 18K11602, 26330190 and 21500161.

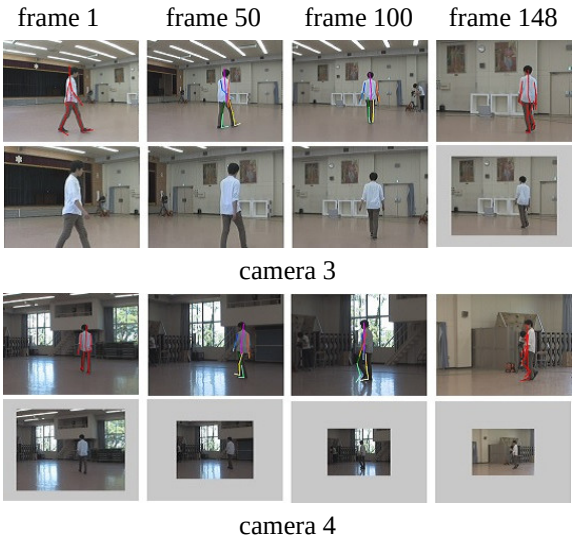

Figure 13: Images taken with the variable focal length and images taken with the focal length fixed at the reference focal length are shown for each camera. The upper row is the images for the variable focal length, where the skeleton of the body is superimposed on the image as a motion capturing result. The lower row is the images for the fixed focal length where the fixed focus image is made from the variable focus image. The 1st, 2nd, 3rd and 4th columns correspond to the 1st, 50th, 100th, and 148th frames, respectively.

REFERENCES

Alcantarilla, P. F., Bartoli, A., and Davison, A. J. (2012). Kaze features. In *ECCV*, volume 31, pages 214–227.

Alcantarilla, P. F., Nuevo, J., and Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale shapes. In *British Machine Vision Conference*, pages 117–126.

Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359.

Cannelle, B., Paparoditis, N., and Tournaire, O. (2010). Panorama-based camera calibration. In *IAPRS*, volume XXXVIII, Part 3, pages 73—78.

Joo, H., Liu, H., Tan, L., Gui, L., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., and Sheikh, Y. (2015). Panoptic studio: A massively multiview system for social motion capture. In *ICCV*.

Kobayashi, D., Igarashi, T., and Yamamoto, M. (2018). Motion capture from multi-view mobile cameras. In *CVIM (Japanese Edition)*, volume 2018-CVIM-210, pages 1–6.

Kobayashi, D. and Yamamoto, M. (2015). Wide-range motion capture from panning multi-view cameras. In *ACM SIGGRAPH Asia 2015 Posters*, page Article 36.

Kobayashi, D. and Yamamoto, M. (2018). Capturing floor exercise from multiple panning-zooming camera. In *Eurographics / ACM SIGGRAPH Symposium on Computer Animation-Posters*.

Kurihara, K., Hoshino, S., Yamane, K., and Nakamura, Y. (2002). Optical motion capture system with pan-tilt camera tracking and realtime data processing. In *International Conference on Robotics & Automation*, pages 1241–1248.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.

Nageli, T., Oberholzer, S., Plüss, S., Alonso-Mora, J., and Hilliges, O. (2018). Flycon: real-time environment-independent multi-view human pose estimation with aerial vehicles. volume 37, page Article 182.

Rhodin, H., Richardt, C., Casas, D., Insafutdinov, E., Shafiei, M., Seidel, H., Schiele, B., and Theobalt, C. (2016). Egocap: Egocentric marker-less motion capture with two fisheye cameras. *ACM Trans. Graph.*, 35(6):Article 162.

Saini, N., Price, E., Tallamraju, R., and Black, M. J. (2019). Markerless outdoor human motion capture using multiple autonomous micro aerial vehicles. In *ICCV*.

Shum, H.-Y. and Szeliski, R. (2000). Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130.

Sinha, S. N. and Pollefeys, M. (2006). Pan?tilt?zoom camera calibration and high-resolution mosaic generation. *Computer Vision and Image Understanding*, 103(3):170–183.

Sundaresan, A. and Chellappa, R. (2005). Markerless motion capture using multiple cameras. In *Computer Vision for Interactive and Intelligent Environment*.

Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3d machine vision. In *CVPR*, pages 364–374.

Ukita, N. and Matsuyama, T. (2005). Real-time cooperative multi-target tracking by communicating active vision agents. *Computer Vision and Image Understanding*, 97(2):137–179.

Wu, Z. and Radke, R. J. (2013). Keeping a pan-tilt-zoom camera calibrated. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1994–2007.

Xu, L., Liu, Y., Cheng, W., Guo, K., Zhou, G., Dai, Q., and Fang, L. (2016). Flycap: Markerless motion capture using multiple autonomous flying cameras. *IEEE Transactions on Visualization and Computer Graphics*, 24(8):2284–2297.

Yamamoto, M. (2005). A simple and robust approach to drift reduction in motion estimation of human body. *The IEICE Transactions on Information and Systems(Japanese Edition)D-II*, J88-D-II(7):1153–1165.

Yamamoto, M., Isono, S., , and Wada, Y. (2014). Determining pose of intertwisting human bodies from multiple camera views. *The journal of the Institute of Image Information and Television Engineers(Japanese Edition)*, 68(8):J358–J370.