

## Gait Recognition from Fisheye Images

Chi Xu Yasushi Makihara Xiang Li Yasushi Yagi  
Osaka University, Osaka, Japan

{xu, makihara, li, yagi}@am.sanken.osaka-u.ac.jp

### Abstract

*Gait recognition has been a hot topic of extensive research in video-based surveillance and forensics. Compared with traditional rectilinear cameras mainly used in existing studies, fisheye cameras have a wider field of view, and hence are more suitable for gait recognition applications in navigation robots, which enables more flexible and free surveillance scenarios. In this paper, to the best of our knowledge, we propose the first framework for gait recognition from images captured by fisheye cameras. To deal with severe image distortion and partial body occlusions induced by fisheye cameras set at lower heights, we combine a set of preprocessing procedures with a state-of-the-art model-based gait recognition method. Specifically, an input fisheye image is first expanded into a panoramic view before pedestrian detection. A person-dependent gnomonic projection is then applied to the detected human region for distortion correction. Next, background regions are attenuated to improve human model fitting accuracy in complex outdoor scenes. The resulting rectified image sequence is finally fed into the gait recognition network for human model estimation and gait feature extraction. To validate the performance, we collect a real fisheye image gait dataset with various views and capture scenarios, including simplified indoor and challenging outdoor scenes. Various experiments on the collected dataset demonstrate the effectiveness of the proposed method.*

### 1. Introduction

Gait recognition, which is a biometric technique that identifies a person based on his/her body shape and posture while walking, has become increasingly popular nowadays. Because of its unique advantages over other biometrics, such as easy capture over a long distance without human cooperation, gait recognition owns great application potential in surveillance, criminal investigation, and forensics [8, 27, 43], and has been extensively studied in the past dozen years [11, 21, 40, 44, 55, 65].

Existing gait recognition studies mainly focus on im-



Figure 1. Gait image examples captured by a fisheye camera. The person circled in red is the walking subject. Because the camera height is set relatively lower, the occlusion increases as he approaches the camera. In addition, there are severe distortions in the fisheye images.

ages captured by traditional rectilinear surveillance cameras, which can only monitor a fixed area with a limited field of view. In contrast, fisheye cameras have an advantage of a wider field of view, which helps to eliminate blind spots in surveillance, and hence, fisheye cameras have recently gained growing attraction in many computer vision tasks, such as person re-identification [4, 14] and pose estimation [63, 68]. Gait recognition using fisheye cameras, which has not been investigated to the best of our knowledge, is also more beneficial for some potential applications. For example, for a gait recognition system mounted on a navigation robot, fewer fisheye cameras are needed to cover a larger space than rectilinear cameras. In this case, the navigation robot can monitor different places more flexibly and freely, making it easier to surveillance tasks such as finding lost children in shopping malls.

However, unlike surveillance cameras set at high positions, the relatively lower height of fisheye cameras mounted on robots (e.g., one-meter height) can easily lead to partial observation of a target person in captured gait video sequences. As shown in Fig. 1, the full body of a person may be observed when he/she is far away from the camera, and the unobserved regions (i.e., out of the image, a kind of partial occlusion) becomes larger and larger as he/she approaches. Most studies on gait recognition for rectilinear images with partial occlusion extract features from appearance-based representations, which require known full-body bounding boxes to keep the input oc-

cluded gait images still size-normalized and human center-registered [28, 48, 50, 52, 62]. In contrast, model-based approaches [37, 39, 66] are believed to have the advantage of addressing occlusions without such a requirement, since a full human model can be fitted even on a partially occluded image with only visible body parts. Therefore, we also deal with gait recognition from fisheye images with occlusions using a model-based method.

Another important issue with fisheye images is the severe distortion due to the lens geometry. As shown in Fig. 1, the distortion of human appearance (e.g., body shape, limb length, and body proportions) changes significantly with the spatial position during walking. Additionally, the orientation of the human body in fisheye images is often inclined, which is another difference from the usual rectilinear images. Because of these challenges, conventional gait recognition methods and even person detection algorithms may easily fail [5, 76] if directly applied to fisheye images.

Therefore, we propose a framework combining a set of preprocessing procedures and a state-of-the-art gait recognition network for gait recognition from fisheye images. The contributions of this work are three-fold.

### **1. First attempt at gait recognition from fisheye images considering distortion and occlusions.**

To the best of our knowledge, this is the first work aimed at gait recognition from images captured by fisheye cameras, taking into account image distortion and partial occlusions, which are easy to occur in real applications, but have been neglected in existing omnidirectional camera-based gait recognition works [56, 57]. Compared with rectilinear cameras, gait recognition using fisheye cameras is more favorable for flexible surveillance applications such as navigation robots owing to the advantage of wider field of views.

### **2. A baseline framework for a reasonable combination of preprocessing and a gait recognition network.**

To mitigate differences between fisheye and normal rectilinear images, we sequentially apply a set of preprocessing procedures for distortion correction and background effect attenuation. In particular, instead of a global distortion correction often used in other computer vision tasks with fisheye lenses [3, 7, 14], we employ a person-dependent projection for the detected human bounding box regions to minimize residual distortion in the corrected human body. The rectified images are then used for gait feature learning through a state-of-the-art occlusion-aware model-based gait recognition network [66], which is properly fine-tuned on the rectified fisheye data.

### **3. Effectiveness validated on real fisheye data.**

We collected the first gait dataset using real fisheye cameras. Multiple cameras were used to capture the subjects' walking sequences from different observation views, such as front and side views. Different capture scenarios

were also considered, including indoor scenes with simplified backgrounds, and outdoor scenes with complex backgrounds and illumination conditions. The effectiveness of the proposed method is validated on the collected dataset through a variety of experimental analyses.

## **2. Related Work**

### **2.1. Gait recognition**

While a few works use omnidirectional cameras without considering distortion and occlusion [56, 57], most gait recognition studies use rectilinear cameras, which can be roughly divided into two categories, appearance-based and model-based methods. Appearance-based methods typically extract features from silhouette-based gait representations, such as gait energy image (GEI) [21] and frequency domain features [44]. Various metric learning techniques [19, 34, 42, 45, 46, 49] and deep networks [22, 36, 55, 60, 65, 67, 71] are further applied to improve feature invariance against various covariates (e.g., views and clothing). Recent CNN-based works explore inputting silhouette sequences by regarding input as an unordered set [11, 12, 24], or incorporating part-based features [18, 25] and 3D CNNs to model gait temporal information [26, 40], which achieve superior performance over GEI-based methods. A few studies directly handle RGB images, excluding useless color information through disentangled representation learning [74, 75] or trainable synthetic silhouettes generation [38].

Model-based approaches often exploit pose features by fitting a human model on the input RGB image. By applying pose estimation models such as OpenPose [10] and HRNet [58], skeletal joint locations are obtained for further feature extraction via CNN [2, 39] or graph convolutional networks [61, 64]. A model-based method was proposed in [37], using both shape and pose features obtained from the skinned multi-person linear (SMPL) [41] model estimated by human mesh recovery (HMR) network [30].

To address occlusion in captured gait videos, existing works mainly use appearance-based methods, which directly learn relatively invariant features from occluded sequences [9, 13, 28, 50, 52, 72], or reconstruct non-occluded silhouettes prior to feature extraction [23, 48, 54, 62]. As a prerequisite, a full-body bounding box including unobserved parts needs to be known to maintain the consistency of human body scales and positions in the input occluded image sequence. A recent occlusion-aware model-based method (OA-ModelGait) [66] works without the above requirement by directly fitting an SMPL model on the occluded image, achieving state-of-the-art recognition performance. Therefore, we employ OA-ModelGait [66] as the gait feature extractor by fine-tuning it on the rectilinear-like images obtained after preprocessing the raw fisheye images.

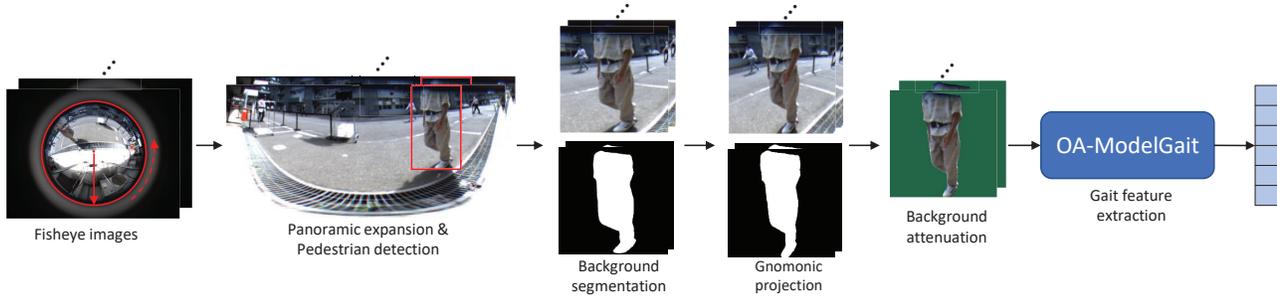


Figure 2. Overview of the proposed method. Given a fisheye image sequence, a set of preprocessing procedures are first applied to rectify distortions and reduce background effects. The resulting images are finally fed into a state-of-the-art gait recognition network, OA-ModelGait, for feature extraction. Only the walking course area is shown in the expanded panorama image.

## 2.2. Other vision tasks using fisheye images

Fisheye cameras have been widely explored in many computer vision tasks, such as surveillance [1, 31, 32], semantic segmentation [17, 53, 59], pose estimation [3, 47, 63, 68, 73], and person re-identification [4, 7, 14, 15, 35, 76]. To cope with the distortion in fisheye images, panoramic images are generated for simple correction in surveillance works [31, 32]. Camera calibration-based correction [3, 7, 14] and image transformation networks [35, 47] are also often used to obtain rectilinear images prior to the main tasks. Some recent studies work directly on fisheye images, such as segmentation using deformable convolutions [17, 53], and 3D pose estimation combined with back-projection [63, 68, 73], which all require the ground-truth on original fisheye images.

## 3. Proposed method

### 3.1. Overview

Figure 2 shows the overview of the proposed method, which includes a set of preprocessing procedures and a gait feature learning network. Given a raw fisheye image, panoramic expansion is first applied to roughly normalize the body orientations of walking subjects. Next, pedestrian detection is performed on the panorama image, followed by instance segmentation for the detected subject. A person-dependent gnomonic projection is then applied to both the obtained RGB and the corresponding silhouette of the bounding box for distortion correction. Finally, background regions are attenuated in RGB to mitigate the effects on subsequent human model fitting. The processed image sequence is input into OA-ModelGait [66] to estimate the SMPL model and extract the gait feature used for final recognition.

### 3.2. Preprocessing

#### 3.2.1 Panoramic expansion

Since existing pedestrian detectors cannot detect subjects with large distortions and inclined bodies from raw fisheye

images, we first apply panoramic expansion based on the longitude-latitude mapping to roughly mitigate image distortion and normalize body orientations to normal upright poses. Let  $P(\varphi, \omega)$  be a point in the longitude-latitude representation on the hemispherical surface and  $p(\rho, \alpha)$  be a point in polar coordinate representation on the fisheye image projected from point  $P$  (see Fig. 3). Assuming that the fisheye lens uses an equidistant mapping function [6], its maximum angle of view is  $\pi$ , and the radius of the fisheye image is  $R$ , we can get the focal length  $f = \frac{2R}{\pi}$ , and the distance from point  $p$  to the fisheye image center  $\rho = f\theta$ , where  $\theta$  is the angle between point  $P$  and the optical axis. Because we have the relationship between the longitude-latitude representation and the polar coordinate as  $\theta = \frac{\pi}{2} - \omega$  and  $\alpha = \varphi + \pi$ , the point  $p(u, v)$  on the image plane can be represented as

$$\begin{aligned} u &= u_0 + \rho \cos \alpha = u_0 + R \left(1 - \frac{2\omega}{\pi}\right) \cos(\varphi + \pi) \\ v &= v_0 + \rho \sin \alpha = v_0 + R \left(1 - \frac{2\omega}{\pi}\right) \sin(\varphi + \pi), \end{aligned} \quad (1)$$

where  $(u_0, v_0)$  is the image coordinate of the center of the fisheye circle. Finally, given  $\omega \in [0, \frac{\pi}{2}]$  and  $\varphi \in [0, 2\pi]$ , the panoramic expansion image is easily obtained by interpolation from the fisheye image as shown in Fig. 3 (b).

#### 3.2.2 Pedestrian detection and background segmentation

We then apply a well-established pedestrian detector, i.e., YOLOv5 [29] to the expanded panorama images, because it is one of the fastest detectors. After localizing the target subject, we apply the instance segmentation method PointRend [33] to obtain the silhouettes for the purpose of the background attenuation described later. Because tight bounding boxes often lead to worse segmentation results for the subjects, we add a margin area of 50% of the bounding box both horizontally and vertically to include a little more background. The loose bounding box area is cropped as input to PointRend, and after segmentation is complete, only

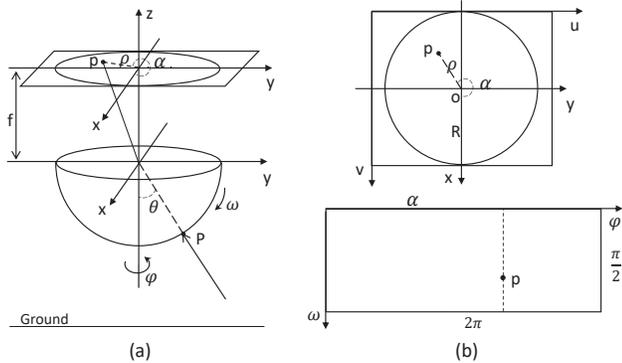


Figure 3. Illustration of panoramic expansion: a) projection from hemispherical surface to image plane; (b) longitude-latitude expansion.

the tight bounding box area is re-cropped for the following processes.

### 3.2.3 Gnomonic projection

Although the panoramic expansion corrects the body orientation, making the detection and segmentation successful, some detailed parts still have large distortion, especially the legs near the bottom of the panorama (see Fig. 2). There are some deep learning-based studies [69, 70] to rectify distortion by training with pre-prepared undistorted ground-truths. To avoid the high cost of the ground-truth preparation, we therefore choose a traditional geometric projection for further distortion correction. Unlike the global projections often used in other works [3, 7], we apply a person-dependent gnomonic projection [16] to detected bounding box area and its corresponding silhouette image to specifically rectify distortions in the human body. Gnomonic projection is a nonconformal map projection that projects points on the surface of a sphere from the sphere’s center onto a tangent plane. As shown in Fig. 4, a point  $P(\varphi, \omega)$  in the longitude-latitude representation on the unit hemispherical surface is projected to a point  $Q(u, v)$  on the image plane, which is parallel to the tangent plane whose tangent point is  $S(\frac{\pi}{2}, 0)$ . In a 3D Cartesian coordinate system, let  $Q = [u_0 - u, d, v_0 - v]^T$ , we have  $P = [(u_0 - u)/D, d/D, (v_0 - v)/D]^T$ , where  $D = \sqrt{(u_0 - u)^2 + d^2 + (v_0 - v)^2}$ ,  $d$  is the distance between the projected image plane and the  $z$ -axis, and  $(u_0, v_0)$  is the image coordinate of the center of the projected image plane. As  $P = [\cos \omega \cos \varphi, \cos \omega \sin \varphi, \sin \omega]^T$ , the longitude and latitude can be computed as

$$\varphi = \arctan \frac{d}{u_0 - u}, \quad \omega = \arcsin \frac{v_0 - v}{D}. \quad (2)$$

Therefore, given  $d$  and  $(u_0, v_0)$  computed based on the target bounding box, the projected image can be easily

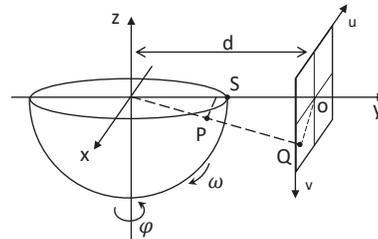


Figure 4. Illustration of gnomonic projection.

obtained by interpolation from the longitude-latitude expanded image.

The projection above is based on the tangent point  $S(\frac{\pi}{2}, 0)$ . In our implementation, we set the tangent point  $S'(\varphi', \omega')$  at the center of the subjects’ bounding box; therefore, we first rotate the point  $P$  around the  $z$ -axis by the angle  $\varphi' - \frac{\pi}{2}$ , and then rotate around the  $x$ -axis by the angle  $\omega'$ .

### 3.2.4 Background attenuation

Due to the complex background and illumination conditions in captured images, especially in outdoor scenes, it becomes very difficult to fit human body models directly on RGB images, and hence, we apply background attenuation to reduce the effect. Given the cropped RGB and corresponding silhouettes after gnomonic projection, we keep the foreground pixels and fill the background pixels with green (see Fig. 2). The obtained images are used as input to the subsequent recognition network.

## 3.3. Gait recognition

### 3.3.1 OA-ModelGait

OA-ModelGait [66] is a state-of-the-art model-based gait recognition method that handles occlusions by directly fitting SMPL models on input images with only visible body parts, unlike other methods that require full-body bounding boxes. OA-ModelGait contains three modules: sequence encoder, occlusion attenuation module, and recognition module. The sequence encoder first estimates 85D SMPL parameters from each input image, which consist of 10D shape, 69D pose, 3D global rotation, and 3D camera parameters. The occlusion attenuation module then reduces the dependence of the estimated SMPL parameters on the input occlusion patterns.

Specifically, the gait phase is first estimated for each input image, and then the phases are synchronized between different input sequences by interpolating the estimated SMPL parameters into the phases of a canonical gait cycle. The synchronized SMPL body parameters are finally transformed into occlusion-independent parameters by using the camera parameters as cues for the input occlusion pattern. The transformed SMPL parameters are then fed

into the recognition module to extract gait features: a shape feature is a shape parameter averaged over a sequence; and a pose feature is learned from the 3D joint locations computed from the SMPL meshes using a CNN.

### 3.3.2 Finetune from pre-trained model

The original OA-ModelGait [66] was pre-trained on synthetic occlusion images from OU-MVLP dataset [51], which was captured by common rectilinear cameras. Considering the difference from rectified fisheye images with real occlusions, we fine-tune the pre-trained model on our data to boost the performance. As for the loss functions used for SMPL estimation, we modify one defined in the original paper [66] by excluding the ground-truth supervision, which is unavailable for our dataset as described below, and re-define the sum of the remaining loss functions as  $L_{\text{other}}$ .

In the pre-training of OA-ModelGait, phase estimation is supervised by the ground-truth phase labels, and constrained by a smoothness loss and a penalty loss to keep the temporal continuity of the estimated phases in a sequence. However, the ground-truth phase labels are difficult to obtain for real fisheye images. Given that the estimation error of the pre-trained model is small for key gait phases (e.g., single-support and double-support phases) of fisheye image sequences, we only use smoothness and penalty losses as the phase loss  $L_{\text{phase}}$  to alleviate the phase estimation errors of other intermediate phases in the fine-tuning stage.

The final estimated SMPL parameters and 3D body joint locations are supervised by the corresponding ground-truth in the pre-training. However, unlike occlusion images synthesized from OU-MVLP’s full-body images, the ground-truth SMPL models are unavailable for our data. Therefore, we use the SMPL body parameters obtained from the pre-trained model as the pseudo ground-truth to compute the estimation loss  $L_{\text{gt}}$  for coarse supervision, and refine the estimation through other losses and subsequent recognition loss. As for the camera parameters (i.e., scale and translation parameters for 2D projection on the original image), we use constraints on the foot joint locations for implicit supervision because we assume the foot bottom is visible and located at the bottom of the cropped image unlike the head top is sometimes unobserved. Specifically, the location of a joint projected onto the image plane can be computed using the camera parameters as [30]

$$c_y = ((p_y + t_y)s + 1)H/2, \quad (3)$$

where  $c_y$  is the vertical image coordinate of a joint,  $H$  is the image height,  $p_y$  is the projected coordinate of the 3D joint on the template 2D image via orthographic projection, and  $s$  and  $t_y$  are the scale and vertical translation parameters, respectively. Ideally, at least one foot should be at the bottom

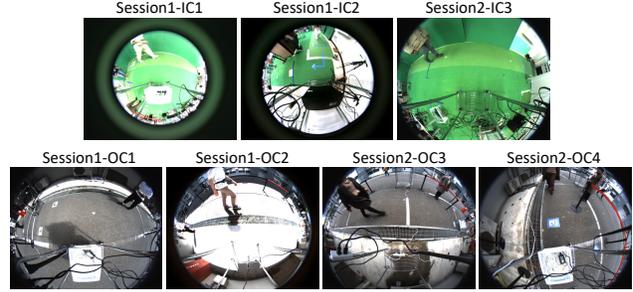


Figure 5. Examples images from the fisheye gait dataset. The text above each image is the label of the camera. “I” indicates indoor, “O” indicates outdoor, and “C” indicates camera. The indoor images have simplified green backgrounds, while the outdoor images have more complex backgrounds and illumination changes. The same camera label between two sessions indicates a similar capture background, but not the same camera/lens model. More examples are shown in the supplementary.

of the image, i.e.,  $c_y^{\text{foot}} = H$ , where  $c_y^{\text{foot}}$  is the larger vertical coordinate between the two foot joints. Therefore, the constraint to minimize the distance between the foot joint location and the image bottom can be defined as

$$L_{\text{foot}} = \|(p_y^{\text{foot}} + t_y)s - 1\|_2^2, \quad (4)$$

which implicitly constrains the scale and translation parameters.

For the final recognition module, considering the contrastive loss [20] is more suitable for verification tasks often used in real applications [27], we use it to optimize the recognition performance, which is computed as

$$L_{\text{recog}} = \frac{1}{N_{\text{pair}}} \sum_{n=1}^{N_{\text{pair}}} (\gamma_n d_n^2 + (1 - \gamma_n) \max(m - d_n^2, 0)), \quad (5)$$

where  $N_{\text{pair}}$  is the number of sample pairs in a mini-batch,  $d_n$  is the dissimilarity of the  $n$ -th pair,  $m$  is the margin, and  $\gamma_n$  is the flag of the  $n$ -th pair, defined as 1 for the same subject pairs and 0 for different subject pairs.

Finally, the overall loss function is defined as the combination of the above losses as

$$L_{\text{all}} = w_{\text{other}}L_{\text{other}} + w_{\text{phase}}L_{\text{phase}} + w_{\text{gt}}L_{\text{gt}} + w_{\text{foot}}L_{\text{foot}} + w_{\text{recog}}L_{\text{recog}}, \quad (6)$$

where  $w_{\text{phase}}$ ,  $w_{\text{sim}}$ ,  $w_{\text{seq}}$ ,  $w_{\text{gt}}$ ,  $w_{\text{foot}}$ , and  $w_{\text{recog}}$  are the weights for each loss.

## 4. Data collection

### 4.1. Capture setup

To validate the performance of the proposed method, we collected a real fisheye gait dataset containing images cap-

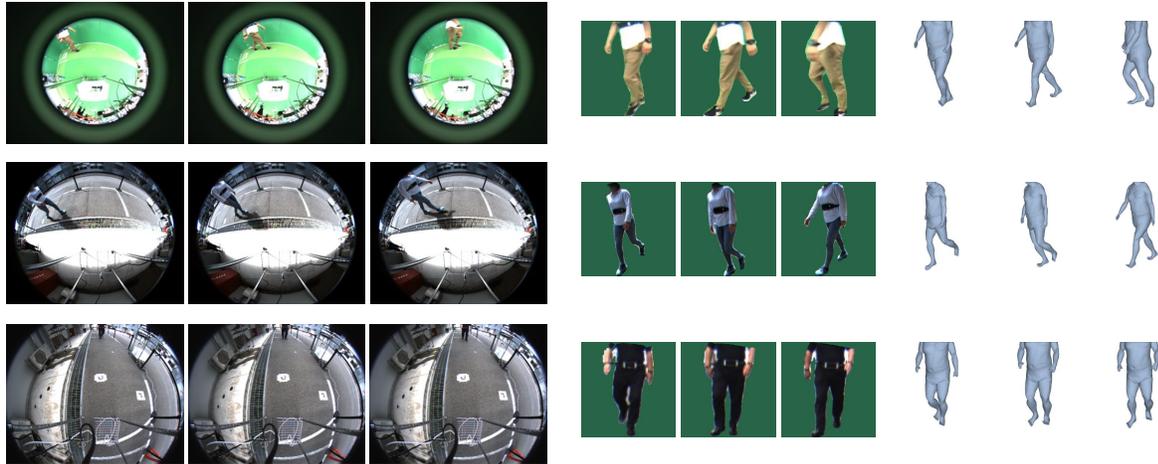


Figure 6. Examples of the preprocessing results and the corresponding SMPL models estimated by the proposed method. The fisheye image samples from the top to bottom are captured by the cameras IC1, OC2, and OC4, respectively.

tured in different scenes. Gait videos at 30 fps were captured by ground-facing fisheye cameras placed on carts approximately one meter high, which simulated the height of cameras mounted on robots. Different cameras (e.g., Point Grey 2.8MP USB camera and HDMABEL Super 2.3Meg camera) and Fujinon FE185 fisheye lenses with different focal lengths (e.g., 1.4 and 2.7 millimeters) were used, thus capturing images at different resolutions (e.g.,  $1280 \times 1024$  and  $1920 \times 1200$ ). Each subject walked twice from the start to the end of the 10-meter-walking course, and hence, two gait videos can be obtained for each subject. A walking course was set up indoors and outdoors, respectively. The former course was simplified with green walls and carpets as the background, and the latter with the campus natural environment as different backgrounds (see Fig. 5). While the subject was walking, multiple cameras captured gait videos from different views, such as front and side views (i.e., angles between the optical axis of the camera and the walking course).

The dataset contains data collected in two sessions conducted in September and December 2021, respectively. In the first session, each of the 176 subjects walked along both the indoor and outdoor walking courses. Two cameras were set up for the indoor course to capture from the front view and the side view near the starting point, respectively. Four cameras were set up for the outdoor course, and the capturing views included front view and side views from both sides of the course. The subjects were separated into two groups, one group participated in the capture in the morning and the other group in the afternoon to include videos under different illumination conditions (see bottom images in Fig. 5). In the second session, walking videos of another 34 subjects were all collected in the afternoon. The cameras were set up in similar positions to the first session, with a slight shift, and an additional camera was added to capture

the indoor course from another side view near the ending point (see the image "Session2-IC3" in Fig. 5).

## 4.2. Protocols

We only use video clips of subjects approaching the cameras (i.e., walking from the start to the camera) as gait sequences for recognition. As a result, most cameras capture sequences containing more than 60 frames, while a few cameras have sequences of less than 20 frames. For each subject, the sequence of the first walk is used as the gallery and the second walk as the probe. We consider two evaluation scenarios: the same-session test, and the across-session test. In the first scenario, 130 subjects from the first session are used to fine-tune the OA-ModelGait model, and other disjoint 46 subjects are used for testing. Due to the illumination difference between images captured in the morning and afternoon, we equally choose subjects from both groups to constitute the training and testing sets. In the second evaluation scenario, all 34 subjects from the second session are used for testing to evaluate the effect of slight changes in camera settings and the natural environment on the performance of the model trained on the first session data.

## 5. Experiments

### 5.1. Implementation details

During preprocessing, we filtered out other irrelevant subjects in the captured image by setting a region of interest for the walking course in the panorama image after applying pedestrian detection. After preprocessing, following [66], we resized the preprocessed images to  $224 \times 224$  and randomly chose 25 consecutive frames from a sequence for training. If a sequence has less than 25 frames, we repeated frames from the beginning. For the fine-tuning of OA-ModelGait, we set the initial learning rate to  $10^{-5}$  for

the first 10K iterations and decreased it by 0.1 for the last 10K iterations. Adam was chosen as the optimizer and the batch size was set to  $8 \times 8$ , indicating that there were eight subjects in a mini-batch, and each subject had eight samples. The margin  $m$  in Eq. (5) was set to 0.2. The weight parameter  $w_{\text{foot}}$  in Eq. (6) was set to 1 and the others set to the same as ones defined in [66]. In the inference phase, if there were more than 25 frames in a sequence, we first used the first 25 frames as input and then applied a sliding window strategy with a step of five frames to iterate over all frames. The final feature of this sequence was the mean of features of all the 25-frame input sub-sequences. As for the evaluation metrics, rank-1 identification rate and equal error rate (EER) were used.

## 5.2. Visualization examples

We choose one indoor camera and two outdoor cameras (i.e., IC1, OC2, and OC4) to visualize the results of pre-processing and SMPL model estimation in Fig. 6. It is obvious that preprocessing successfully mitigates effects such as distortion in fisheye images, generating images similar to normal rectilinear images for both indoor and outdoor cameras. SMPL model fitting by fine-tuned OA-ModelGait also works well on preprocessed images, estimating human models with similar shapes and poses to the input image. Although the SMPL models estimated from the images of the front-view camera OC4 are not completely facing forward due to estimation errors in the 3D global rotation parameters, body shapes and poses are still consistent with the subjects in the image, which are actually used for recognition.

## 5.3. Comparison with state-of-the-arts

To demonstrate the effectiveness of the proposed method in gait recognition, we compare with three state-of-the-art gait recognition methods, i.e., GaitSet [11], GaitGL [40], and ModelGait [37]. Similar preprocessing procedures are applied to the comparison methods to generate corresponding inputs. Specifically, since GaitSet and GaitGL are silhouette-based methods, we use the silhouettes obtained after gnomonic projection as inputs; ModelGait is also an SMPL model-based method that works on RGB images, and thus we use the same inputs as ours. For a fair comparison, we also fine-tune the models of the three methods pre-trained on synthetic occlusion images from OUMVLP [51]. Additionally, following [37, 66], we train separate models for shape and pose features, and apply score-level fusion to obtain the final results for ModelGait and the proposed method.

The results of all comparison methods on the fisheye gait dataset are shown in Table 1. For both the same-session test and cross-session test, the proposed method achieves the best rank-1 identification rate and EER, demonstrating

Table 1. Rank-1 identification rate (%) and EER (%) of each comparison method in the same-session and cross-session tests on the fisheye gait dataset. The mean results for all six/seven camera combinations in the probe and gallery are reported. Bold and bold italic indicate the best and second-best results.

Methods	Same-session		Cross-session	
	Rank-1	EER	Rank-1	EER
GaitSet [11]	70.3	8.11	57.4	12.97
GaitGL [40]	61.8	15.28	50.9	15.59
ModelGait [37]	<b>89.8</b>	<b>2.35</b>	<b>79.3</b>	<b>4.81</b>
Ours	<b>92.6</b>	<b>2.00</b>	<b>83.5</b>	<b>3.37</b>

that the chosen OA-ModelGait, which is designed to handle occlusion in rectilinear images, is also more effective than other methods on occluded fisheye images. Compared with model-based methods (i.e., ModelGait and our method) that directly fit human models on input images, appearance-based features are more susceptible to segmentation errors caused by complex outdoor backgrounds and fisheye lens distortions that still somewhat exist even after preprocessing, which leads to the worse performance of GaitSet and GaitGL. For the cross-session test, all methods exhibit performance degradation due to changes in the outdoor natural environment (e.g., changes in background scenes and illumination) and the subject seasonal clothing changes (i.e., late summer vs. winter).

We also report the results of the proposed method for each individual combination of cameras in the same-session and cross-session tests in Tables 2 and 3. Essentially, matching results within the same camera are much better than across different cameras. However, due to view variations between two sequences even with the same camera (e.g., camera OC1)<sup>1</sup>, and unstable walking including acceleration phases in sequences that are too short (e.g., IC1), the same-camera matching results are also somewhat affected in some cases.

The cross-camera matching results are worse because different cameras may contain variations in camera and lens models, scenes (e.g., indoor vs. outdoor), illumination, and walking directions (e.g., side vs. front), which have a greater impact on cross-session test since no corresponding training samples are included. Therefore, matching between cameras with certain similarities yields relatively better performance (e.g., OC2 vs. OC3 in Table 2 with similar capture scene and walk direction, IC1 vs. OC1 in Table 3 with similar camera model).

## 5.4. Ablation study

We finally analyze the effect of preprocessing procedures of the proposed method. Because pedestrian detection does not work without panoramic expansion, we only

<sup>1</sup>This is because the subjects may walk along both sides of the walking course during the two walks, and some cameras are very close to the walking course (e.g., less than 1 meter).

Table 2. Rank-1 identification rate (%) and EER (%) of the proposed method for each individual combination of six cameras in the same-session test. Probe and gallery are denoted by P and G, respectively.

P \ G	IC1	IC2	OC1	OC2	OC3	OC4
IC1	97.8 / 0.48	82.6 / 3.29	89.1 / 2.17	84.8 / 2.17	89.1 / 2.17	80.4 / 3.19
IC2	93.5 / 2.17	100 / 0.00	91.3 / 2.17	95.7 / 2.17	97.8 / 1.30	95.7 / 3.72
OC1	89.1 / 4.35	84.8 / 2.17	100 / 0.19	95.7 / 2.17	95.7 / 2.17	95.7 / 2.17
OC2	91.3 / 2.27	89.1 / 1.98	97.8 / 1.74	100 / 0.48	97.8 / 0.29	87.0 / 3.04
OC3	91.3 / 2.17	95.7 / 0.87	97.8 / 0.72	95.7 / 2.17	100 / 0.00	97.8 / 3.00
OC4	82.6 / 2.56	87.0 / 2.46	89.1 / 2.17	84.8 / 4.35	91.3 / 3.29	100 / 0.10

Table 3. Rank-1 identification rate (%) and EER (%) of the proposed method for each individual combination of seven cameras in the cross-session test.

P \ G	IC1	IC2	IC3	OC1	OC2	OC3	OC4
IC1	100 / 1.43	64.7 / 5.88	79.4 / 3.39	91.2 / 2.94	52.9 / 5.88	70.6 / 5.44	52.9 / 7.13
IC2	70.6 / 4.63	100 / 0.18	91.2 / 1.34	67.6 / 4.81	88.2 / 2.94	88.2 / 2.50	76.5 / 3.65
IC3	82.4 / 4.46	94.1 / 1.34	100 / 0.00	73.5 / 4.19	97.1 / 2.94	100 / 0.62	94.1 / 2.94
OC1	94.1 / 2.14	58.8 / 8.11	70.6 / 6.33	97.1 / 2.14	50.0 / 7.58	64.7 / 6.95	61.8 / 5.88
OC2	70.6 / 4.99	94.1 / 2.32	97.1 / 1.34	76.5 / 5.88	100 / 0.00	100 / 0.18	97.1 / 2.94
OC3	73.5 / 3.39	91.2 / 3.12	97.1 / 1.25	73.5 / 5.88	100 / 0.18	100 / 0.00	94.1 / 2.67
OC4	58.8 / 6.95	82.4 / 3.65	94.1 / 2.94	73.5 / 6.33	91.2 / 1.87	94.1 / 1.16	100 / 0.27

Table 4. Rank-1 identification rate (%) and EER (%) for the same-session and cross-session tests for ablation study. The mean results for all six/seven camera combinations in the probe and gallery are reported. "w/o Seg." and "w/o Proj." indicate no background attenuation and gnomonic projection, respectively.

Methods	Same-session		Cross-session	
	Rank-1	EER	Rank-1	EER
Our w/o Seg.	83.6	3.81	71.4	6.15
Our w/o Proj.	84.7	3.64	79.6	4.89
Ours	<b>92.6</b>	<b>2.00</b>	<b>83.5</b>	<b>3.37</b>

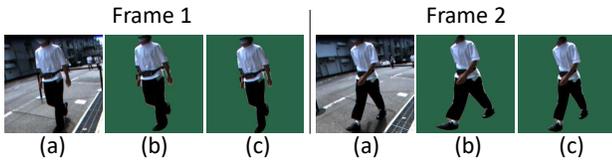


Figure 7. Results of two example frames after different preprocessing. (a) and (b) are the processed images without background attenuation and gnomonic projection, respectively. (c) is the image with the full preprocessing procedures of the proposed methods.

focus on background attenuation and gnomonic projection. Specifically, we first remove the process of background attenuation, and hence, the input images for the recognition network retain the original backgrounds, including buildings and possibly irrelevant people (see (a) in Fig. 7). We then remove the gnomonic projection from the preprocessing, which results in still large distortions in the input images, especially for the lower body parts (see (b) in Fig. 7). According to the recognition results shown in Table 4, removing either of these two procedures degrades

the performance, demonstrating their necessity in the proposed method. Compared with removing gnomonic projection, the method without background attenuation performs worse, because the complex backgrounds affect the accuracy of model fitting more. Another reason is that the pre-training of OA-ModelGait is only performed on images with green backgrounds, which also limits the fine-tuned performance on complex backgrounds.

## 6. Conclusion

We propose the first framework for gait recognition from fisheye images, which is more suitable for flexible gait recognition applications. Given a fisheye image, a set of preprocessing procedures are sequentially applied to normalize body orientation, correct distortion, and mitigate background effects. The processed images are then fed into a state-of-the-art occlusion-aware gait recognition network to estimate SMPL models and extract gait features for recognition. Experiments on the collected fisheye gait dataset validate the effectiveness of the proposed method.

Currently, we have collected fisheye gait data from about 200 subjects. An important future work is to collect data from more subjects and more diverse capture scenes, which is especially needed for outdoor cases. Since the current framework is a combination of separate preprocessing and gait recognition, another future direction is to investigate a unified framework that directly handles fisheye images, which may improve recognition performance thanks to end-to-end training.

## References

- [1] Raghad Al-Harasis and Belal H. Sababha. On the design and implementation of a dual fisheye camera-based surveillance vision system. *Multimedia Tools and Applications*, 78:22667–22689, 2019. 3
- [2] Weizhi An, Shiqi Yu, Yasushi Makihara, Xinhui Wu, Chi Xu, Yang Yu, Rijun Liao, and Yasushi Yagi. Performance evaluation of model-based gait on multi-view very large population database with pose sequences. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):421–430, 2020. 2
- [3] Kohei Aso, Dong-Hyun Hwang, and Hideki Koike. Portable 3d human pose estimation for human-human interaction using a chest-mounted fisheye camera. In *Augmented Humans Conference 2021*, page 116–120, New York, NY, USA, 2021. Association for Computing Machinery. 2, 3, 4
- [4] Arko Barman, Wencheng Wu, Robert P. Loce, and Aaron Michael Burry. Person re-identification using overhead view fisheye lens cameras. In *2018 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–7, 2018. 1, 3
- [5] Massimo Bertozzi, Luca Castangia, Stefano Cattani, Antonio Prioletti, and Pietro Versari. 360 detection and tracking algorithm of both pedestrian and vehicle using fisheye images. In *2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 132–137, 2015. 2
- [6] Felix Bettonvil. Fisheye lenses. *WGN, Journal of the International Meteor Organization*, vol. 33, no. 1, p. 9-14, 33:9–14, 2005. 3
- [7] Gregor Blott, J. Yu, and Christian Heipke. Multi-view person re-identification in a fisheye camera network with different viewing directions. *PGF – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 87:263–274, 2019. 2, 3, 4
- [8] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011. 1
- [9] Nikolaos V. Boulgouris and Zhiwei X. Chi. Human gait recognition based on matching of body components. *Pattern Recognition*, 40(6):1763–1770, 2007. 2
- [10] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2
- [11] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proc. of the 33th AAAI Conference on Artificial Intelligence (AAAI 2019)*, 2019. 1, 2, 7
- [12] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3467–3478, 2022. 2
- [13] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11):977–984, 2009. 2
- [14] Mertcan Cokbas, John Bolognino, Janusz Konrad, and Prakash Ishwar. Frida: Fisheye re-identification dataset with annotations. In *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8, 2022. 1, 2, 3
- [15] Mertcan Cokbas, Prakash Ishwar, and Janusz Konrad. Spatio-visual fusion-based person re-identification for overhead fisheye images. *ArXiv*, abs/2212.11477, 2022. 3
- [16] H. S. M. Coxeter. *Introduction to Geometry, 2nd Edition*, pages 93, 289–290. New York: Wiley, 1969. 4
- [17] Liuyuan Deng, Ming Yang, Hao Li, Tianyi Li, Bing Hu, and Chunxiang Wang. Restricted deformable convolution-based road scene semantic segmentation using surround view cameras. *IEEE Transactions on Intelligent Transportation Systems*, 21(10):4350–4362, 2020. 3
- [18] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. Gaitpart: Temporal part-based model for gait recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14213–14221, 2020. 2
- [19] Y. Guan, C. Li, and F. Roli. On reducing the effect of covariate factors in gait recognition: A classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(7):1521–1528, July 2015. 2
- [20] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742, June 2006. 5
- [21] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006. 1, 2
- [22] Y. He, J. Zhang, H. Shan, and L. Wang. Multi-task gans for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security*, 14(1):102–113, Jan 2019. 2
- [23] Martin Hofmann, Daniel Wolf, and Gerhard Rigoll. Identification and reconstruction of complete gait cycles for person identification in crowded scenes. In *VISAPP 2011 - Proceedings of the International Conference on Computer Vision Theory and Application*, pages 594–597, 01 2011. 2
- [24] Saihui Hou, Chunshui Cao, Xu Liu, and Yongzhen Huang. Gait lateral network: Learning discriminative and compact representations for gait recognition. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX*, page 382–398, Berlin, Heidelberg, 2020. Springer-Verlag. 2
- [25] Xiaohu Huang, Duowang Zhu, Hao Wang, Xinggang Wang, Bo Yang, Botao He, Wenyu Liu, and Bin Feng. Context-sensitive temporal feature learning for gait recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12889–12898, 2021. 2
- [26] Zhen Huang, Dixiu Xue, Xu Shen, Xinmei Tian, Houqiang Li, Jianqiang Huang, and Xian-Sheng Hua. 3d local convolutional neural networks for gait recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14900–14909, 2021. 2

- [27] H Iwama, D. Muramatsu, Y. Makihara, and Y. Yagi. Gait verification system for criminal investigation. *IPSN Transactions on Computer Vision and Applications*, 5:163–175, Oct. 2013. 1, 5
- [28] Yumi Iwashita, Koji Uchino, and Ryo Kurazume. Gait-based person identification robust to changes in appearance. *Sensors*, 13(6):7884–7901, 2013. 2
- [29] Glenn Jocher. Yolov5. <https://github.com/ultralytics/yolov5>, 2020. 3
- [30] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 5
- [31] Hyungtae Kim, Eunjung Chae, Gwanghyun Jo, and Joonki Paik. Fisheye lens-based surveillance camera for wide field-of-view monitoring. In *2015 IEEE International Conference on Consumer Electronics (ICCE)*, pages 505–506, 2015. 3
- [32] Hyungtae Kim, Jaehoon Jung, and Joonki Paik. Fisheye lens camera based surveillance system for wide field of view monitoring. *Optik*, 127(14):5636–5646, 2016. 3
- [33] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808, 2020. 3
- [34] Worapan Kusakunniran, Qiang Wu, Jian Zhang, and Hongdong Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Proc. of IEEE computer society conference on Computer Vision and Pattern Recognition 2010*, pages 1–8, San Francisco, CA, USA, Jun. 2010. 2
- [35] Tianjiao Li, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, and Zhiheng Li. Uav-human: A large benchmark for human behavior understanding with unmanned aerial vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16266–16275, June 2021. 3
- [36] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. Gait recognition via semi-supervised disentangled representation learning to identity and covariate features. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2
- [37] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, Shiqi Yu, and Mingwu Ren. End-to-end model-based gait recognition. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020. 2, 7
- [38] Junhao Liang, Chao Fan, Saihui Hou, Chuanfu Shen, Yongzhen Huang, and Shiqi Yu. Gaitedge: Beyond plain end-to-end gait recognition for better practicality. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, page 375–390, Berlin, Heidelberg, 2022. Springer-Verlag. 2
- [39] Rijun Liao, Shiqi Yu, Weizhi An, and Yongzhen Huang. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98:107069, 2020. 2
- [40] Beibei Lin, Shunli Zhang, and Xin Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14648–14656, October 2021. 1, 2, 7
- [41] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: A skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), Oct. 2015. 2
- [42] Jiwen Lu and Yap-Peng Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382–393, 2010. 2
- [43] Niels Lynnerup and Peter Kastmand Larsen. Gait as evidence. *IET Biometrics*, 3(2):47–54, 6 2014. 1
- [44] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conference on Computer Vision*, pages 151–163, Graz, Austria, May 2006. 1, 2
- [45] Y. Makihara, A. Suzuki, D. Muramatsu, X. Li, and Y. Yagi. Joint intensity and spatial metric learning for robust gait recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6786–6796, July 2017. 2
- [46] Raul Martin-Felez and Tao Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, 47(12):3793 – 3806, 2014. 2
- [47] Daisuke Miki, Shinya Abe, Shi Chen, and Kazuyuki Demachi. Robust human pose estimation from distorted wide-angle images through iterative search of transformation parameters. *Signal, Image and Video Processing*, 14, 06 2020. 3
- [48] Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi. Gait regeneration for recognition. In *2015 International Conference on Biometrics (ICB)*, pages 169–176, 2015. 2
- [49] D. Muramatsu, A. Shiraiishi, Y. Makihara, M. Z. Uddin, and Y. Yagi. Gait-based person recognition using arbitrary view transformation model. *IEEE Transactions on Image Processing*, 24(1):140–154, 2015. 2
- [50] Prasit Nangtin, Pinit Kumhom, and Kosin Chamnongthai. Gait identification with partial occlusion using six modules and consideration of occluded module exclusion. *J. Vis. Commun. Image Represent.*, 36(C):107–121, Apr. 2016. 2
- [51] Takemura Noriko, Makihara Yasushi, Muramatsu Daigo, Echigo Tomio, and Yagi Yasushi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSN Transactions on Computer Vision and Applications*, 10(1):4, Feb 2018. 5, 7
- [52] Javier Ortells, Ramón Alberto Mollineda, Boris Mederos, and Raúl Martín-Félez. Gait recognition from corrupted silhouettes: a robust statistical approach. *Machine Vision and Applications*, 28:15–33, 2017. 2
- [53] Clément Ployat, Ola Ahmad, Freddy Lécué, and Farida Cheriet. Adaptable deformable convolutions for semantic segmentation of fisheye images in autonomous driving systems. *CoRR*, abs/2102.10191, 2021. 3
- [54] Aditi Roy, Shamik Sural, Jayanta Mukherjee, and Gerhard Rigoll. Occlusion detection and gait silhouette reconstruction from degraded scenes. *Signal, Image and Video Processing*, 5:415–430, 2011. 2

- [55] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016. 1, 2
- [56] Kazushige Sugiura, Yasushi Makihara, and Yasushi Yagi. Gait identification based on multi-view observations using omnidirectional camera. In *Proc. 8th Asian Conf. on Computer Vision*, pages 452–461, Tokyo, Japan, Nov. 18–22 2007. LNCS 4843. 2
- [57] Kazushige Sugiura, Yasushi Makihara, and Yasushi Yagi. Omnidirectional gait identification by tilt normalization and azimuth view transformation. In *Proceedings of the IEEE Workshop on OMNIVIS*, Marseille, France, 2008. 2
- [58] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2
- [59] Álvaro Sáez, Luis M. Bergasa, Elena López-Guillén, Eduardo Romera, Miguel Tradacete, Carlos Gómez-Huélamo, and Javier del Egado. Real-time semantic segmentation for fisheye urban driving images based on erfnet. *Sensors*, 19(3), 2019. 3
- [60] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018. 2
- [61] Torben Teepe, Ali Khan, Johannes Gilg, Fabian Herzog, Stefan Hörmann, and Gerhard Rigoll. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2314–2318, 2021. 2
- [62] Md. Zasim Uddin, Daigo Muramatsu, Noriko Takemura, Md Atiqur Rahman Ahad, and Yasushi Yagi. Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. *IPSJ Transactions on Computer Vision and Applications*, 11:1–18, 2019. 2
- [63] Jian Wang, Lingjie Liu, Weipeng Xu, Kripasindhu Sarkar, and Christian Theobalt. Estimating egocentric 3d human pose in global space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11500–11509, October 2021. 1, 3
- [64] Likai Wang, Jinyan Chen, and Yuxin Liu. Frame-level refinement networks for skeleton-based gait recognition. *Computer Vision and Image Understanding*, 222:103500, 2022. 2
- [65] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan. A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, 2017. 1, 2
- [66] Chi Xu, Yasushi Makihara, Xiang Li, and Yasushi Yagi. Occlusion-aware human mesh model-based gait recognition. *IEEE Transactions on Information Forensics and Security*, 18:1309–1321, 2023. 2, 3, 4, 5, 6, 7
- [67] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu. Cross-view gait recognition using pairwise spatial transformer networks. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2020. 2
- [68] Weipeng Xu, Avishek Chatterjee, Michael Zollhöfer, Helge Rhodin, Pascal Fua, Hans-Peter Seidel, and Christian Theobalt. Mo2cap2: Real-time mobile 3d motion capture with a cap-mounted fisheye camera. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2093–2101, 2019. 1, 3
- [69] Shangrong Yang, Chunyu Lin, Kang Liao, Chunjie Zhang, and Yao Zhao. Progressively complementary network for fisheye image rectification using appearance flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6348–6357, 2021. 4
- [70] Xiaoqing Yin, Xinchao Wang, Jun Yu, Maojun Zhang, Pascal Fua, and Dacheng Tao. Fisheyerecnet: A multi-context collaborative deep network for fisheye image rectification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 469–484, 2018. 4
- [71] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 532–539, July 2017. 2
- [72] Shiqi Yu, Daoliang Tan, Kaiqi Huang, and Tieniu Tan. Reducing the effect of noise on human contour in gait recognition. In Seong-Whan Lee and Stan Z. Li, editors, *Advances in Biometrics*, pages 338–346, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. 2
- [73] Yahui Zhang, Shaodi You, Sezer Karaoglu, and Theo Gevers. Multi-person 3d pose estimation from a single image captured by a fisheye camera. *Computer Vision and Image Understanding*, 222:103505, 2022. 3
- [74] Ziyuan Zhang, Luan Tran, Feng Liu, and Xiaoming Liu. On learning disentangled representations for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):345–360, 2022. 2
- [75] Ziyuan Zhang, Luan Tran, Xi Yin, Yousef Atoum, Xiaoming Liu, Jian Wan, and Nanxin Wang. Gait recognition via disentangled representation learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4705–4714, 2019. 2
- [76] Zixuan Zhao, Zihao Zhao, Song Wang, Paul Watta, and Yi Lu Murphey. Pedestrian re-identification using a surround-view fisheye camera system. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021. 2, 3