Omnidirectional Gait Identification by Tilt Normalization and Azimuth View Transformation

Kazushige Sugiura, Yasushi Makihara, and Yasushi Yagi

The Institute of Scientific and Industrial Research, Osaka University, Japan
{sugiura, makihara, yagi}@am.sanken.osaka-u.ac.jp

Abstract. We propose an effective matching method of gait sequences containing changes of observed tilt and azimuth views provided by an omnidirectional camera. Given gait image sequence, observed tilt view and depth to the subject are normalized by re-projecting the images to a virtual tilt-free image plane orthogonally, and a series of tilt- and size-normalized gait silhouette images are constructed as Gait Silhouette Volume (GSV). On the other hand, a relatively wide range of azimuth view changes can be continuously observed by tracking the subject in the omnidirectional image, multi-view frequency-domain features can be obtain from the GSV as attractive multiple cues for identification. Observed azimuth view ranges of two gait sequences do, however, not always coincide, therefore a View Transformation Model (VTM) supported by reference view addition with geometrical model is exploited to synthesize missing-view gait features for each gallery and probe sequence in matching phase. Experiments of gait identification including 22 subjects from different view ranges demonstrate the effectiveness of the proposed method.

1 Introduction

Gait-based person identification is considered as a promising method for various applications such as wide-area surveillance because gait personalities can be observed at a distance from a camera, while other biometrics such as faces and fingerprints needs a person's approach to a sensor.

Recently, many gait identification methods have been studied [1][2][3]. One of the difficulties facing those approaches is an appearance change due to changes of viewing or walking direction. Yu et al. [4] discussed the effects of view angle variation on gait identification and reported a performance drop when view difference is large. This fact implies that gait images from different views contain different types of features.

From this point of view, previous works handling view changes have been divided into two directions: view-invariant approaches and multi-view approaches.

View-invariant approaches try to match different-view gait image sequences by extracting view-invariant features or by transforming features' views. Kale et

al. [5] proposed a method to synthesize arbitrary-view images from a single-view image with perspective projection by assuming a walking person as a planer object parallel to his/her sagittal plane (call it sagittal plane assumption). Makihara et al. [6] proposed a View Transformation Model (VTM) which enables to transform a few-view reference gait features to arbitrary-view features. These methods, however, treat only azimuth view changes, that is, they do not consider tilt view changes.

Multi-view approaches tries to improve identification performance integrating multi-view gait features for matching as multiple cues. Wang et al.[7] introduced Dempster-Shafer (D-S) rule for fusing multi-view gait sequences. Sugiura et al. [8] used an omnidirectional camera to observe a wide range of view transition of a tracked walking person. The methods [7][8] are, however, not effective when an overlapped view range (call it "intersection views") between a gallery and a probe sequences is narrow. For example, if the intersection view is just a single view, the methods are degraded from multi-view approaches to single-view approaches in essentials.

Therefore, we propose an effective matching method of gait sequences containing changes of observed tilt and azimuth views provided by an omnidirectional camera. In the gait sequence captured by the omnidirectional camera, a range of observed azimuth view change is much wider than that of observed tilt view change, hence we make the most of observed azimuth view changes as attractive multiple cues for identification, while we normalize observed tilt change as nuisance for identification. First, because the sagittal plane assumption [5] works well for normalization of relatively small view change, we extend the assumption to the tilt normalization. Second, gait features for the exclusive-or observed azimuth views of gallery and probe views are synthesized by the VTM [6] to make the most of view ranges of a gallery and probe sequence. Then, features both for the intersection views and the exclusive-or views (call joined views "union views") are matched and the matching results are integrated for better identification. Note that the proposed method matches gait features for union views, while the previous methods [7][8] do so only for intersection views.

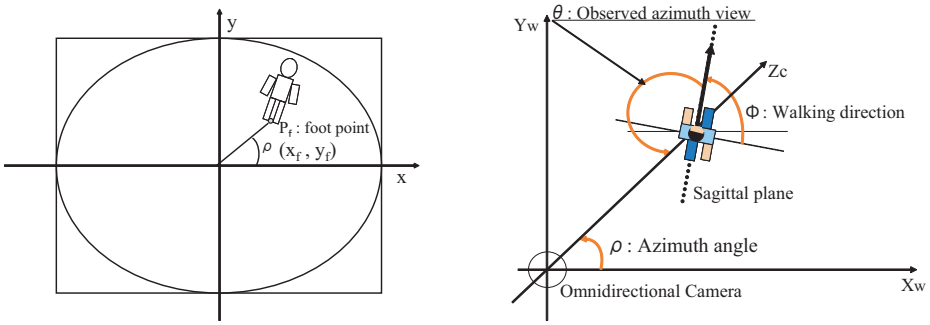
The outline of this paper is as follows. First, construction of GSV considering tilt and depth normalization is addressed in section 2. Second, multi-view gait feature extraction from a omnidirectional image sequence is presented in section 3. Then, matching of gait sequences with different view ranges are addressed with the VTM formulation and reference view addition with geometrical model in section 4. Moreover, experimental results of gait identification are shown in section 5. Section 6 contains conclusions and further works in the area.

2 Construction of tilt and depth normalized GSV

A GSV is composed of spatio-temporal gait silhouette images which are normalized by height, tilt, depth, and the center of the region and which are aligned on the temporal axis. In this paper, the silhouette candidates are extracted by background subtraction and shadow removal, and then graph-cut algorithm segments out the silhouette (see [9] for detail). In the following sections, we describe a procedure for construction of tilt and depth normalized GSV.

2.1 Observed azimuth view estimation

The first step is estimation of observed azimuth view. The observed azimuth view θ is defined as $\theta = (\pi - \phi) + \rho$, where ρ is an azimuth angle, and ϕ is a walking direction (see Fig. 1). In this paper, we calculate these values based on a foot point as shown in Fig. 1 (a).



(a) Omnidirectional image coordinate (b) World coordinate

Fig. 1. Definition of observed azimuth view and sagittal plane

The azimuth angle ρ is simply defined as direction of a foot point $P_f(x_f, y_f)$ in the omnidirectional image as shown in Fig. 3. The walking direction ϕ is estimated from a trajectory of foot points $P_F(X_{Wf}, Y_{Wf}, Z_{Wf})$ in a world coordinate. Here, the world coordinate is defined by an origin which is a middle point between a mirror focal point O_m and camera center O_c (see Fig. 1), and by three axes : X and Y axes which are parallel to x and y axes in the omnidirectional image, and Z axis whose direction is vertical against the ground.

Let (r_f, ρ) and (R_f, ρ) be the foot points of polar coordinate in the omnidirectional image and on the floor plane respectively. If the floor plane is regarded as another image plane, the distance H_r from mirror focal point O_m to the floor plane can be seen as focal length to the floor image plane. Then, radius R_f is calculated as [10]

$$R_f = \frac{-(b^2 - c^2)H_r r_f}{(b^2 + c^2)f - 2bc\sqrt{r_f^2 + f^2}}, \tag{1}$$

where f is focal length for the omnidirectional image plane, and b and c are hyperboloidal mirror parameters. Thus, walking trajectory on the floor plane is obtained as a time series of the foot points (R_f, ρ).

Next, walking direction ϕ is defined as a tangential direction of the estimated walking trajectory. Let $(X_{Wf}(n), Y_{Wf}(n))$ and $(V_{X_{Wf}(n)}, V_{Y_{Wf}(n)})$ be foot point's position and velocity at nth frame, and velocity is introduced by central difference as follows

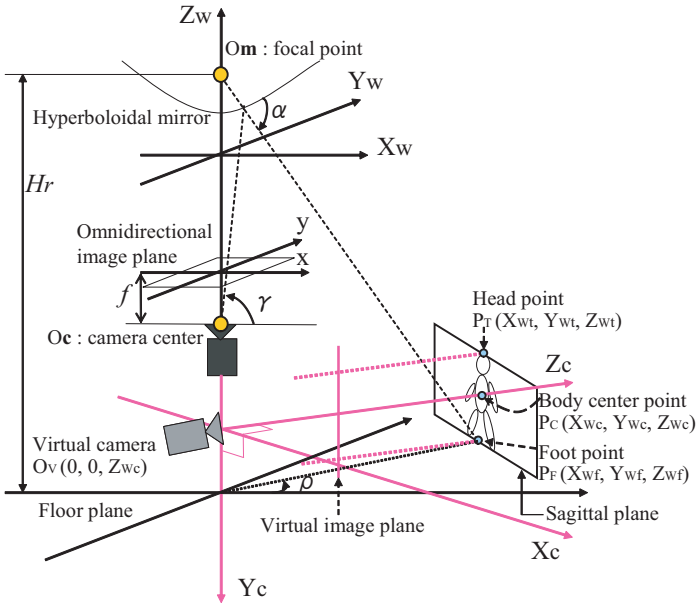


Fig. 2. Relation between sagittal plane and normalized image plane

$$V_{X_{Wf}(n)} = \frac{X_{Wf}(n+\Delta n) - X_{Wf}(n-\Delta n)}{2\Delta n} \tag{2}$$

$$V_{Y_{Wf}(n)} = \frac{Y_{Wf}(n+\Delta n) - Y_{Wf}(n-\Delta n)}{2\Delta n}. \tag{3}$$

Here, Δn is set to be 15 [frame] considering velocity smoothness. Finally, walking direction ϕ_n in nth frame is defined as direction of velocity vector $(V_{X_{Wf}(n)}, V_{Y_{Wf}(n)})$.

Figure 3 shows a transition of the azimuth angle ρ, the walking direction ϕ, and the observed azimuth view θ for a straight-walk sequence. While the walking direction ϕ keeps the almost the same values, the azimuth angle ρ changes widely from 340 deg to 280 deg. As a result, even if the walking trajectory is straight, the wide range of the views can be gained thanks to the azimuth angle changes as shown in Fig. 3.

2.2 Tilt and depth normalization

The second step is GSV construction considering tilt and depth normalization. Because a tilt changes by a depth from a camera to a subject, captured images vary even if observed azimuth view does not change. For example, when the subject walks near the camera, both feet do not appear in the same vertical position on the image due to perspective projection. Therefore tilt and depth normalization is needed. Concretely speaking, we assume a walking subject as a plane object parallel to his/her sagittal plane (call it sagittal plane assumption), and create a silhouette image on the sagittal plane. Then, a virtual camera is set on the same height as the body center point (let it be $P_C(X_{Wc}, Y_{Wc}, Z_{Wc})$),

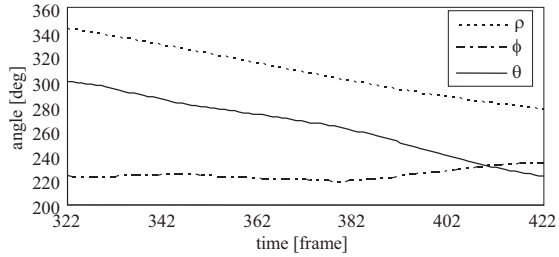


Fig. 3. Changes of azimuth angle ρ, walking direction ϕ, and observed azimuth view θ

and tilt and depth are normalized by orthogonal projection from the silhouette image with the camera. Note that Kale et al. [5] normalize observed azimuth view under the sagittal plane assumption, and that in case of large view difference, transformation error becomes large due to self occlusion (see [6]). For example, in case of front oblique view, backward arm swing for the opposite side to a camera is occluded by the body, while that in case of side view is not occluded. On the other hand, we apply sagittal plane assumption to relatively small change of tilt, hence the self occlusion effects are much less.

Thus, a sagittal plane is defined as the first step to the normalization. It is supposed that the subject walks heading toward walking direction and stands vertical to the floor plane (see Fig. 1). Based on the assumption, the sagittal plane turns out to pass the foot point P_F and is composed by two vectors of vertical direction and walking direction, and it is defined as

$$\tan\phi = \frac{Y_W - Y_{Wf}}{X_W - X_{Wf}}. \tag{4}$$

Next, let the head point in the sagittal plane and the omnidirectional image be $P_T(X_{Wt}, Y_{Wt}, Z_{Wt})$ and $P_t(x_t, y_t, z_t)$ respectively. Then a ray of the head point in the world coordinate is defined as

$$\tan\rho = Y_W / X_W = y_t / x_t \tag{5}$$

$$Z_W = \sqrt{1 + \tan^2} X_W \tan\alpha + c, \tag{6}$$

where, γ and α are elevation and depression angles defined in Fig. 2 respectively, and they are defined as

$$\alpha = \tan^{-1} \frac{(b^2 + c^2)\sin\gamma - 2bc}{(b^2 - c^2)\cos\gamma} \tag{7}$$

$$\gamma = \tan^{-1} \frac{f}{\sqrt{x_t^2 + y_t^2}}. \tag{8}$$

As a result, the head point P_T is solved by Eqs. from (4) to (8). Finally, a body center point P_C is given as a middle point between the head point P_T and foot point P_F as follows

$$X_{Wc} = \frac{X_{Wt} + X_{Wf}}{2}, Y_{Wc} = \frac{Y_{Wt} + Y_{Wf}}{2}, Z_{Wc} = \frac{Z_{Wt} + Z_{Wf}}{2}. \tag{9}$$

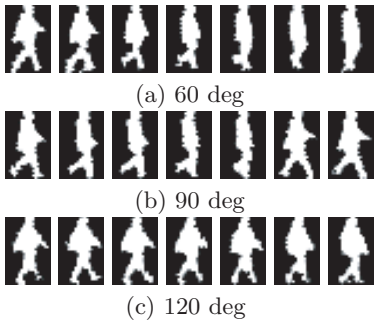
Next, each point silhouette in the omnidirectional image coordinate is projected to the sagittal plane in the same way as the head point, and it subsequently transformed into a point in the virtual camera coordinate. Here, the virtual camera coordinate is defined by an origin $O_V(0, 0, Z_{Wc})$, and Z-axis which passes body center and is parallel to the floor plane, Y-axis which is perpendicular downward, and X-axis which is orthogonal to Y and Z axes (see Fig. 2). Let R and $\boldsymbol{t}$ be rotation matrix and translation vector from the world coordinate to the virtual camera coordinate. They are defined as

$$R = \begin{bmatrix} \cos(\frac{\pi}{2} - \rho) & 0 & \sin(\frac{\pi}{2} - \rho) \\ \sin(\frac{\pi}{2} - \rho) & 0 & -\cos(\frac{\pi}{2} - \rho) \\ 0 & -1 & 0 \end{bmatrix} \tag{10}$$

$$\boldsymbol{t} = \begin{bmatrix} 0 & 0 & -Z_{Wc} \end{bmatrix}^T . \tag{11}$$

Then, each point defined in the virtual camera coordinate is projected to a virtual image plane orthogonally. Finally, the image is normalized so that the height $(Z_{Wt} - Z_{Wf})$ can be just predefined height H_g, and so that the aspect ratio of image can be kept. Here, H_g is set to be 30 pixel empirically. Finally a GSV is constructed by aligning images on the temporal axis.

Figure 4 shows GSV examples for multiple observed azimuth views. We can clearly see appearance changes in each view, while both feet lie on almost the same vertical position on the image thanks to tilt and depth normalization.



(a) 60 deg

(b) 90 deg

(c) 120 deg

Fig. 4. GSV examples for multiple observed azimuth views

3 Multi-view gait feature extraction

In this section, Multi-view gait feature extraction is described briefly (see [8] for detail).

3.1 Frequency-domain feature extraction

First, gait period N_{gait} is detected by maximizing the normalized autocorrelation of the constructed GSV for the temporal axis. Next, subsequence $\boldsymbol{S}_{n_s}$ composed of N_{gait} frames is picked up from a complete sequence $\boldsymbol{S}$. A Discrete Fourier Transformation (DFT) for the temporal axis is then applied for the subsequence,

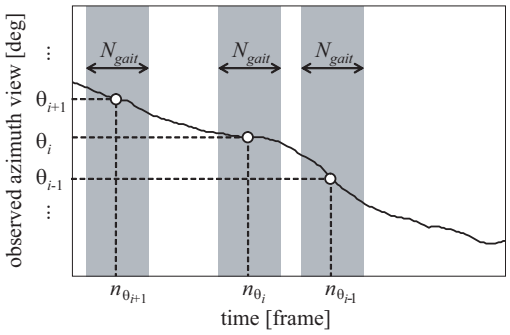


Fig. 5. Overview of multi-view feature extraction

and amplitude spectra normalized with gait period N_{gait} are calculated. In this paper, direct-current elements (0-times frequency, averaged silhouette) and low-frequency elements (1- and 2-times frequencies) are chosen as gait features. Let $\boldsymbol{a}$ be a feature vector composed of elements of the amplitude spectra. As a results, the dimension of the feature vector $\boldsymbol{a}$ sums up to $20 \times 30 \times 3 = 1800$.

3.2 Multi-view feature extraction

In this section, multi-view feature extraction is introduced based on the estimated observed azimuth views. First, multiple basis views $\theta_i (i = 1, 2, \ldots)$ are chosen from observed azimuth views at 15 deg intervals. Next, a basis frame n_{θ_i} corresponding to a basis view θ_i is found from a complete sequence, and a subsequence is picked up as a set of N_{gait} frames around the basis frame n_{θ_i} as shown in Fig. 5.

Results of multi-view feature extraction for multiple subjects are shown in Fig. 6. In this figure, each block indicates each subject, and each row and column indicate observed azimuth view and frequency respectively. We can see individual differences, for example, swing motion difference of subject 2 and 4 in 2-times frequency of 270-deg features. In addition, we can also see view differences for each subject. Thus, by integrating the different type of features across views, gait identification performance should improve more than the case of a single-view feature. Next section gives how to match the multi-view features.

4 Gait feature matching using a VTM

4.1 Overview of VTM

First, a typical implementation of a VTM is shown in Fig. 7. In the training phase, gait features are collected from a limited number of training subjects from all views to form a training set that is then used to create the VTM. Gait features from identification-target subjects are then stored from only a few reference views. Note that the identification-target subjects are different from the training subjects for the VTM and that not all-view galleries but few-view galleries are stored for the identification-target subjects. In the identification phase, this reference set is transformed into the gallery set with the same view as the probe view using the VTM and then matching between the gallery and the probe is processed.

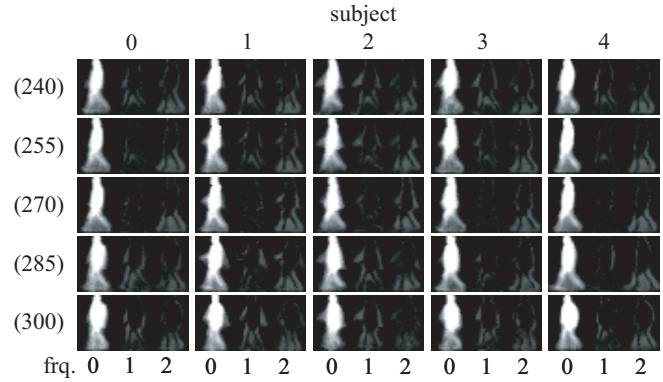


Fig. 6. Multi-view features for each subject (every 15 deg)

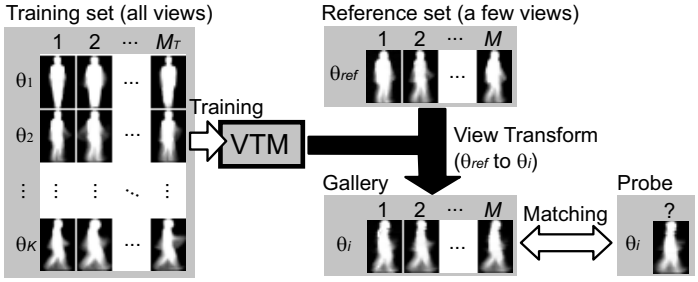


Fig. 7. Problem setting of VTM

4.2 Formulation of VTM

Views are first quantized into K views. Let $\boldsymbol{a}_{\theta_k}^m$ be an N_A dimensional feature vector for the kth view of the mth training subject. Supposing that the feature vectors for K views of M subjects are obtained as a training set, a matrix can be constructed whose row indicates view changes and whose column indicates individuals; and so can be decomposed by Singular Value Decomposition (SVD) as

$$\begin{bmatrix} \boldsymbol{a}_{\theta_1}^1 \cdots \boldsymbol{a}_{\theta_1}^M \\ \vdots \ \ddots \ \vdots \\ \boldsymbol{a}_{\theta_K}^1 \cdots \boldsymbol{a}_{\theta_K}^M \end{bmatrix} = USV^T = \begin{bmatrix} P_{\theta_1} \\ \vdots \\ P_{\theta_K} \end{bmatrix} \left[\boldsymbol{v}^1 \cdots \boldsymbol{v}^M \right], \tag{12}$$

where U is the $KN_A \times M$ orthogonal matrix, V is the $M \times M$ orthogonal matrix, S is the $M \times M$ diagonal matrix composed of singular values, P_{θ_k} is the $N_A \times M$ submatrix of US, and $\boldsymbol{v}^m$ is the M dimensional column vector.

The vector $\boldsymbol{v}^m$ is an intrinsic feature vector of the mth subject and is common for all views. The submatrix P_{θ_k} is a projection matrix from the intrinsic vector $\boldsymbol{v}$ to the feature vector for view θ_k, and is common for all subjects. Thus, the feature vector $\boldsymbol{a}_{\theta_i}^m$ for the view θ_i of the mth subject is represented as

$$\boldsymbol{a}_{\theta_i}^m = P_{\theta_i} \boldsymbol{v}^m. \tag{13}$$

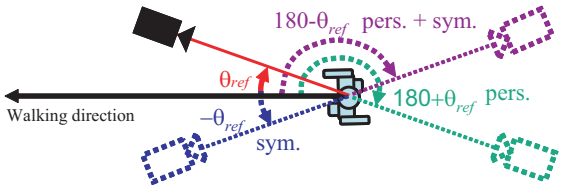


Fig. 8. Overview of reference view addition

Feature vector transformation from N_θ^{ref} reference views (let them be $\boldsymbol{\theta}^{ref} = \{\theta_{ref}(k)|k = 1, \cdots, N_\theta^{ref}\}$ is now easily obtained by least-square method as

$$\hat{\boldsymbol{a}}_{\theta_i}^m = P_{\theta_i} \begin{bmatrix} P_{\theta^{ref}(1)} \\ \vdots \\ P_{\theta^{ref}(N_\theta^{ref})} \end{bmatrix}^+ \begin{bmatrix} \boldsymbol{a}_{\theta^{ref}(1)}^m \\ \vdots \\ \boldsymbol{a}_{\theta^{ref}(N_\theta^{ref})}^m \end{bmatrix}, \tag{14}$$

where $[P_{\theta^{ref}(1)}^T, \cdots, P_{\theta^{ref}(N_\theta^{ref})}^T]^{T+}$ is the pseudo inverse matrix of $[P_{\theta^{ref}(1)}^T, \cdots, P_{\theta^{ref}(N_\theta^{ref})}^T]^T$. Note that this transformation can be applied to an identification-target subject as well as the training subjects because the transformation matrix $P_{\theta_i}[P_{\theta^{ref}(1)}^T, \cdots, P_{\theta^{ref}(N_\theta^{ref})}^T]^{T+}$ does not include terms related to subjects.

In summary, given a few-view references $\{\boldsymbol{a}_{\theta^{ref}(1)}^m \cdots \boldsymbol{a}_{\theta^{ref}(N_\theta^{ref})}^m\}$ are given for an identification-target subject, arbitrary-view galleries $\hat{\boldsymbol{a}}_{\theta_i}^m$ can be transformed from the references by the VTM.

4.3 Reference view addition with geometrical model

In this section, we briefly describe addition of virtual reference view based on the geometrical model to make view transformation more precise (see [11] for detail). First, when weak perspective projection is assumed (call it pers.), the silhouette image observed from the opposite side $(\theta + 180)$ deg becomes a mirror image of the original silhouette image thanks to tilt and depth normalization. Moreover, when a left-right symmetry of gait motion is assumed (call it sym.), the silhouette image from $(360 - \theta)$ deg also becomes a mirror image of the original. When both of the weak perspective projection and the symmetry are assumed (call it pers.+sym.), the silhouette image from $(180 - \theta)$ deg becomes the same image of the original. Figure. 8 shows overview of these assumptions.

Hence, once a gait feature for reference view θ_{ref} is obtained, we can virtually add the same feature for view $(180 - \theta_{ref})$ and a mirror feature for view $(\theta_{ref} + 180)$ and $(360 - \theta_{ref})$.

4.4 Gait feature transformation for exclusive views

Let S^P and S^G be complete sequences for probe and galley respectively. Also observed azimuth views for probe and gallery are defined as $\boldsymbol{\theta}^P = \{\theta^P(i)|i =$

$1, \cdots, N_\theta^P\}$ and $\boldsymbol{\theta}^G = \{\theta^G(j)|j = 1, \cdots, N_\theta^G\}$, and corresponding gait features as $\boldsymbol{a}^P_{\theta^P(i)}$ and $\boldsymbol{a}^G_{\theta^G(j)}$ respectively. Then, an intersection and union set between probe and gallery are defined views as intersection views $\boldsymbol{\theta}^{int} = \{\theta^{int}(k)|k = 1, \cdots, N_\theta^{int}\}$ and union views $\boldsymbol{\theta}^{uni} = \{\theta^{uni}(k)|k = 1, \cdots, N_\theta^{uni}\}$ respectively. In addition, a set of excluded views (missing views) for probe and gallery are defined as excluded views $\boldsymbol{\theta}^{P,ex} = \{\theta^{P,ex}(i)|i = 1, \cdots, N_\theta^{P,ex}\}$ and $\boldsymbol{\theta}^{G,ex} = \{\theta^{G,ex}(j)|j = 1, \cdots, N_\theta^{G,ex}\}$ respectively. Note that excluded views for gallery and probe are discarded in the matching procedure in the previous works such as [8].

Next, probe gait features for the excluded views $\boldsymbol{\theta}^{P,ex}$ are recovered in order of the following priorities.

1. Reference addition with geometrical model from the observed azimuth views $\boldsymbol{\theta}^P$ (let them be $\boldsymbol{\theta}^{P,add}$).
2. VTM from the observed azimuth views $\boldsymbol{\theta}^P$ and the added views $\boldsymbol{\theta}^{P,add}$.

In the same way, gallery gait features for $\boldsymbol{\theta}^{G,ex}$ are transformed.

As a result, we can define probe and gallery gait features for the union views as $\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)}$ and $\boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}$ by using original and transformed gait features respectively.

4.5 Matching

In this section, matching of the union-view gait features is addressed. The matching measure for each union view is simply chosen as the Euclidean distance as

$$d(\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)}, \boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}) = \|\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)} - \boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}\|. \tag{15}$$

Next, matching measures between two complete sequences S^P and S^G are defined. Because the distance distributions are different for each view, z-normalization is processed before integration as

$$d_z(\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)}, \boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}) = \frac{d(\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)}, \boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}) - \overline{d_{\theta^{uni}(k)}}}{\sigma_{d_{\theta_{uni}(k)}}}, \tag{16}$$

where $\overline{d_{\theta^{uni}(k)}}$ and $\sigma_{d_{\theta^{uni}(i)}}$ are an average and standard deviation of distances of all subjects in the gallery for each union view $\theta^{uni}(k)$ respectively, and d_z is a normalized distance. Finally, an average of the normalized distance is defined as complete matching measure:

$$D(S^P, S^G) = \frac{1}{N_\theta^{uni}} \sum_{k=1}^{N_\theta^{uni}} d_z(\boldsymbol{a}^{P,uni}_{\theta^{uni}(k)}, \boldsymbol{a}^{G,uni}_{\theta^{uni}(k)}). \tag{17}$$

5 Experiments

5.1 Datasets

A dataset of walking sequences was obtained in a room (see Fig. 9) and a total of 22 subjects joined the experiments (see Tab. 1 for a breakdown list of the

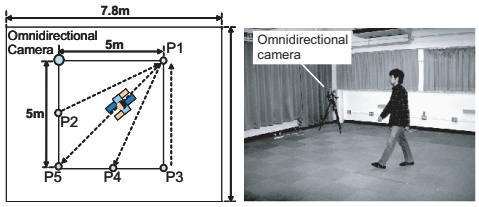


Fig. 9. Experimental setup (left) and scene (right)

Table 1. Breakdown list of subjects

Age	Male	Female	Total
20-29	10	4	14
30-39	5	2	7
40-49	0	0	0
50-59	0	1	1
Total	15	7	22

subjects). The camera used was Sony Inc. DCR-VX2000, and images were captured by 720 × 480 pixel size at 30 fps. The hyperboloidal mirror and camera parameters were $a = 13.722$, $b = 11.708$, $c = 18.038$, $f = 427.944$ (unit: mm). The omnidirectional camera was placed at an diagonal point of 5m by 5m square space and each subject was asked to walk between specified points as shown in Fig. 9. Totally 2 test sets composed of 4 sequences were obtained for each subject and multiple basis views were extracted at 15 deg intervals for each sequence as shown in Tab. 2. Though the proposed method can also treat curving sequences, the sequences in this experiment were limited to straight one to easily obtain almost the same walking courses for all the subjects.

5.2 Feature transformation

The dataset is divided into 2 sets: training set for the VTM and test set for gait identification experiment including 11 subjects respectively, namely, there are no overlapped subjects between training and test set.

In the previous methods (e.g. [8]), only a single intersection view $\{75\}$ and two intersection views $\{255, 270\}$ are used for matching in experiment set Ex1 and Ex2 respectively. On the other hand, transformed features for excluded views are also used in the proposed method as shown in Fig. 10.

Table 2. Gait sequences and their views (G: Gallery, P: Probe)

Set	Sequence	Course	Views
Ex1	S1 (G)	$P3$-$P1$	60, 75
	S2 (P)	$P2$-$P1$	75, 90, 105, 120
Ex2	S3 (G)	$P1$-$P4$	255, 270
	S4 (P)	$P1$-$P5$	240, 255, 270, 285, 300

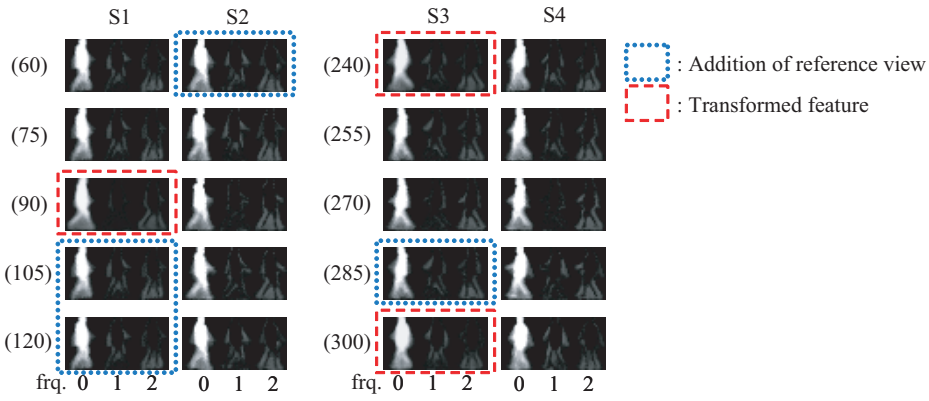


Fig. 10. Feature transformation

For experiment set Ex1, excluded views for sequence S1 (Gallery) are $\{90, 105, 120\}$, and virtual reference views $\{105, 120\}$ are added from 2 gallery reference views $\{75, 60\}$ respectively, then feature for view $\{90\}$ is transformed from 4 gallery reference views $\{60, 75, 105, 120\}$. On the other hand, an excluded view for sequence S2 (Probe) is $\{60\}$, and virtual reference view $\{60\}$ is added from a gallery reference view $\{120\}$. As a result, a total of 5 pairs of gait features for views $\{60, 75, 90, 105, 120\}$ are obtained, and they are used for matching.

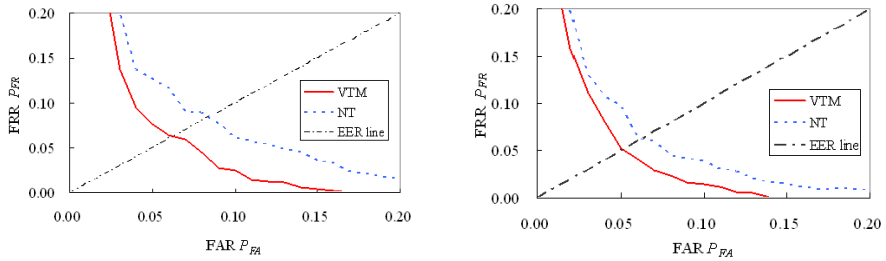
For experiment set Ex2, because views $\{255, 270\}$ for sequence S3 (gallery) are included in views $\{240, 255, 270, 285, 300\}$ for sequence S4 (probe), no features are added to S4. Then virtual reference view $\{285\}$ are added from a gallery reference view $\{255\}$, and then missing views for sequence S3 $\{240, 300\}$ are transformed from 3 gallery reference views $\{255, 270, 285\}$. As a result, a total of 5 pairs of gait features are obtained.

5.3 Gait identification results

The proposed method was compared with the multiple intersection view method [8] to confirm the effectiveness of feature transformation. In this section, the proposed method and the previous method are denoted by "VTM" and "NT (No Transformation)" respectively.

The gait identification experiments were repeated 20 times for different combination of training and test sets. The results were evaluated by Receiver Operating Characteristics (ROC) curves [12]. The ROC curves shows relation between False Rejection Rate (FRR) P_{FR} and False Acceptance Rate (FAR) P_{FA} when the receiver changes the acceptance thresholds. In the ROC evaluation, a curve near the left bottom corner indicates high performance because it realizes both low false rejection and low false acceptance.

ROC curves for experiment sets in Tab. 2 are illustrated in Fig. 11. As a whole, ROC curves for VTM are closer to the left bottom than those for NT. Then, Equal Error Rate (EER) is picked up by calculating an intersection of the ROC curve and EER line $P_{FR} = P_{FA}$ as a typical performance measure for ROC curve.



(a) Ex1 (b) Ex2

Fig. 11. ROC curves

As a result, EERs decrease from 8.4% for NT to 6.3% for VTM in Ex1, 6.0% for NT to 5.0% for VTM in Ex2 respectively, and it turns out the effectiveness of the proposed method.

6 Conclusion and future works

We propose an effective matching method of gait sequences containing changes of observed tilt and azimuth views provided by an omnidirectional camera.

Given gait image sequence, observed tilt view and depth to the subject are normalized by re-projecting the images to a virtual tilt-free image plane orthogonally under a assumption that a walking subject is a plane object parallel to his/her sagittal plane [5]. Then, a series of tilt- and size-normalized gait silhouette images are constructed as Gait Silhouette Volume (GSV) and frequency-domain feature of the GSV are calculated subsequently. Because a relatively wide range of azimuth view changes can be continuously observed by tracking the subject in the omnidirectional image, multi-view features can be obtain from a sequence as attractive multiple cues for identification. Observed azimuth view ranges of the probe and the gallery do, however, not always coincide, therefore a View Transformation Model (VTM) [6] supported by reference view addition with geometrical model is exploited to synthesize excluded-view (missing-view) gait features for each gallery and probe sequence in matching phase. Note that the proposed method exploited union views between the probe and gallery to make the most of observed azimuth view ranges while previous works such as [8] exploited gait features only for intersection views, that is, discarded the excluded views.

Gait identification experiments including 22 subjects from different view ranges were made to show the effectiveness of the proposed method. Compared with the previous work [8], Equal Error Rate (EER) decreases from 8.4% to 6.3% for sequences with a single intersection view, and does from 6.0% to 5.0% for those with a two intersection views. Consequently, it was clear that the proposed method contributed to better performance.

In this paper, though equal weights are used when integrating matching score for all the union views for simplicity, matching reliance is actually dependent on the following factors:

– Whether features for matching are original or transformed ones, namely, whether a view of the feature is intersection view or excluded view between a probe and gallery.
– View transformation accuracy derived from the number of reference views or from view differences between a transformed view and the reference views.

Hence, one of future directions is the optimal weights assignment in matching considering the above factors. We propose an effective matching method of gait sequences containing changes of observed tilt and azimuth views provided by an omnidirectional camera.

This research was supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (S).

References

1. Sarkar, S., Phillips, J., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: Data sets, performance, and analysis. Trans. of Pattern Analysis and Machine Intelligence **27** (2005) 162–177
2. Nixon, M., Carter, J.: Automatic recognition by gait. Proc. of the IEEE **94** (2006) 2013–2024
3. Han, J., Bhanu, B.: Individual recognition using gait energy image. Trans. on Pattern Analysis and Machine Intelligence **28** (2006) 316– 322
4. Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: Proc. of 7th Asian Conf. on Computer Vision. Volume 1. (2006) 807–816
5. Kale, A., Roy-Chowdhury, A., Chellappa, R.: Towards a view invariant gait recognition algorithm. In: Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance. (2003) 143–150
6. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Proc. of the 9th European Conf. on Computer Vision. Volume 3., Graz, Austria (2006) 151–163
7. Wang, Y., Yu, S., Wang, Y., Tan, T.: Gait recognition based on fusion of multi-view gait sequences. In: Proc. of the IAPR Int. Conf. on Biometrics 2006. (2006) 605–611
8. Sugiura, K., Makihara, Y., Yagi, Y.: Gait identification based on multi-view observations using omnidirectional camera. In: Proc. 8th Asian Conference on Computer Vision, Tokyo, Japan (2007) 452–461 LNCS 4843.
9. Makihara, Y., Yagi, Y.: Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In: Proc. of the 19th Int. Conf. on Pattern Recognition. (2008) (to appear).
10. Yamazawa, K., Yagi, Y., Yachida, M.: Hyperomni vision: Visual navigation with an omnidirectional image sensor. Systems and Computers in Japan **28** (1997) 36–47
11. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Which reference view is effective for gait identification using a view transformation model? In: Proc. of the IEEE Computer Society Workshop on Biometrics 2006, New York, USA (2006)
12. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The feret evaluation methodology for face-recognition algorithms. Trans. of Pattern Analysis and Machine Intelligence **22** (2000) 1090–1104