Gait Identification Based on Multi-view Observations Using Omnidirectional Camera

Kazushige Sugiura, Yasushi Makihara, and Yasushi Yagi

Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, Japan
{sugiura,makihara,yagi}@am.sanken.osaka-u.ac.jp

Abstract. We propose a method of gait identification based on multi-view gait images using an omnidirectional camera. We first transform omnidirectional silhouette images into panoramic ones and obtain a spatio-temporal Gait Silhouette Volume (GSV). Next, we extract frequency-domain features by Fourier analysis based on gait periods estimated by autocorrelation of the GSVs. Because the omnidirectional camera makes it possible to observe a straight-walking person from various views, multi-view features can be extracted from the GSVs composed of multi-view images. In an identification phase, distance between a probe and a gallery feature of the same view is calculated, and then these for all views are integrated for matching. Experiments of gait identification including 15 subjects from 5 views demonstrate the effectiveness of the proposed method.

1 Introduction

There is a growing necessity in modern society to identify individuals in many situations, including, surveillance and access control. For personal identification, many biometrics-based authentication methods are proposed using a wide variety of cues; fingerprint, iris, face, and gait. Among these, gait identification has recently gained considerable attention because gait promises to enable surveillance systems to ascertain identity at a distance.

Currently, many gait identification approaches are proposed by model base [1][2] and appearance base [3][4]. One of the difficulties facing those approches is an appearance change due to changes of viewing or walking direction. Yu et al. [5] discussed the effects of view angle variation on gait identification and reported a performance drop when view difference is large.

To cope with view changes, Kale et al. [6] proposed a view transformation method based on perspective projection of the sagittal plane. The method does not, however, work well when view difference is large. Shakhnarovich et al. [7] proposed a visual hull-based method. However, the method needs multiple-view synchronized images for all subjects.

As a training-based method, View Transformation Model (VTM) in the frequency domain was proposed [8]. Once the VTM is trained using sets of gait features of multiple views and subjects, a few-view reference can be transformed

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 452–461, 2007.

into an arbitrary-view gallery so as to match a probe view. They also reported that verification rate increased as the number of reference views increased [9]. Moreover, a method of multi-view gait identification using walking direction changes in a sequence was proposed [10], and it was reported that the verification rate increased as the number of walking directions increased.

It is, however, troublesome to capture gait images many times to acquire many references in registration phase. In addition, it is unreasonable to assume that subjects always change their walking directions enough for multi-view identification.

Therefore, we propose a method of gait identification based on multi-view observations from an omnidirectional camera. Note that an omnidirectional camera makes it possible to observe multi-view gait images even if a subject walks straight. Observation views are estimated by azimuth angles of tracked person regions in the omnidirectional image and walking trajectory on the floor. Then, for each gallery and probe sequence, a silhouette-based gait features are extracted for multiple basis views which are common both for the gallery and the probe. Finally the extracted multiple gait features are matched for each same view and the matching results are integrated for better identification.

The outline of this paper is as follows. First, construction of a Gait Silhouette Volume (GSV) is addressed with silhouette extraction and panoramic expansion in section 2. Next, extraction and matching of multi-view frequency-domain gait feature are described in section 3. Finally, experimental results for gait identification are presented with an analysis of the effect of view variations in section 4. Section 5 contains conclusions and discussion of further work in the area.

2 GSV Construction

2.1 Extraction of Gait Silhouette Images

The first step in constructing a GSV is to extract gait silhouette images by background subtraction from omnidirectional images.

First, background is modeled by average color vector $\overline{\boldsymbol{u}}(x,y)$ and its covariance matrix $\boldsymbol{\Sigma}(x,y)$ at position (x,y) using background image sequence as follows.

$$\overline{\boldsymbol{u}}(x,y) = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{u}(x,y,n) \tag{1}$$

$$\boldsymbol{\Sigma}(x,y) = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{u}(x,y,n)\boldsymbol{u}(x,y,n)^T - \overline{\boldsymbol{u}}(x,y)\overline{\boldsymbol{u}}(x,y)^T, \tag{2}$$

where $\boldsymbol{u}(x,y,n)$ is a background color vector at position (x,y) at nth frame, and N is the number of total frames for a background training sequence.

Second, to extract foreground regions, Mahalanobis distance $D(x,y,n)$ between an input image $\boldsymbol{c}(x,y,n)$ and the modeled background is calculated at each position (x,y) at each nth frame as

$$\boldsymbol{d}(x,y,n) = \boldsymbol{c}(x,y,n) - \overline{\boldsymbol{u}}(x,y) \tag{3}$$

$$D(x,y,n) = \sqrt{\boldsymbol{d}(x,y,n)^T \boldsymbol{\Sigma}(x,y)^{-1} \boldsymbol{d}(x,y,n)}. \tag{4}$$

A foreground region is defined as a set of pixels whose Mahalanobis distance $D(x, y, n)$ is larger than threshold value D_{thresh}. Here, the threshold D_{thresh} is set to be 12.0 empirically. Figure 1 shows an input image and a result of background subtraction. We can see that person region is extracted correctly.

Background subtraction, however, sometimes fails because of cast shadows and illumination condition changes. To overcome such difficulties, a shadow removal is processed based on color vector angle between background and foreground. Moreover, morphological closing filter is applied to improve silhouette quality.

2.2 Panorama Extension

The second step is panorama extension of silhouettes in omnidirectional image [11]. Let $P(X, Y, Z)$ be a point in world coordinate and $p(x, y)$ be a point in an omnidirectional image projected from point P. Then, let ρ and Z be azimuth angle and vertical position in a cylindrical coordinate whose center axis passes through mirror focal point O_m and camera center O_c, and whose radius is R_P. Thus, the panorama extension is expressed as follows.

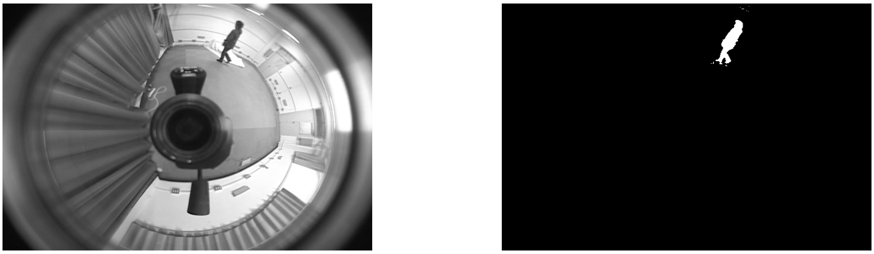
$$\tan\rho = Y/X = y/x \tag{5}$$

$$Z = R_P \tan\alpha + c \tag{6}$$

where $\alpha = \tan^{-1}\frac{(b^2+c^2)\sin\gamma - 2bc}{(b^2-c^2)\cos\gamma}$ and $\gamma = \tan^{-1}\frac{f}{\sqrt{x^2+y^2}}$ are viewing directions defined in Fig. 2 respectively, and b and c are mirror parameters.

2.3 Scaling and Registration of Silhouette Images

The third step is scaling and registration of the panoramic silhouettes to acquire normalized gait patterns.



(a) Input image with omnidirectional camera

(b) Background subtraction

Fig. 1. Result of background subtraction

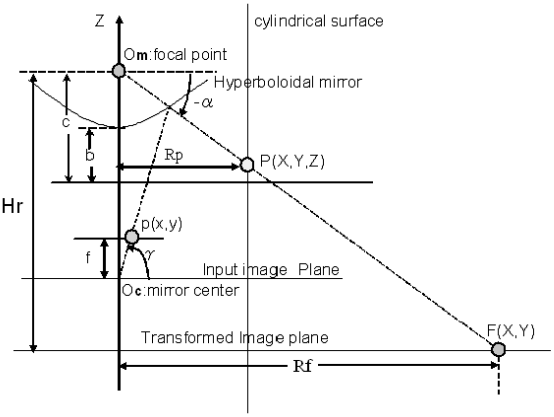
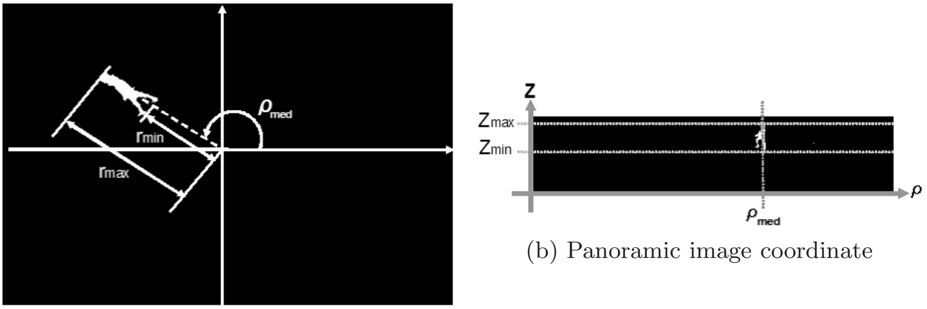


Fig. 2. Projection to cylindrical surface and floor surface



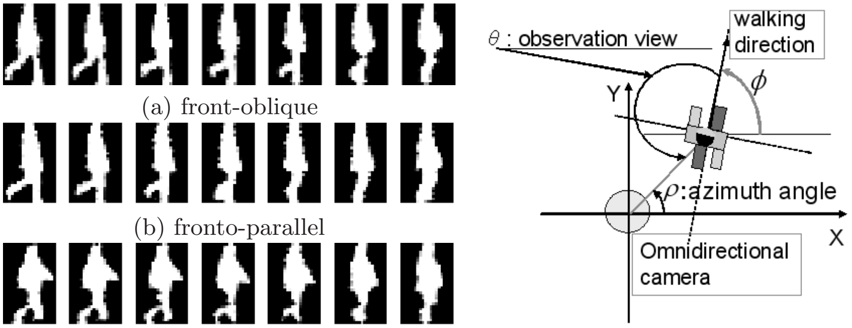
(a) Omnidirectional image coordinate

(b) Panoramic image coordinate

Fig. 3. Definition of person region for scaling and registration

First, person regions are simply tracked in the omnidirectional image considering connected region's area sizes, and position differences between adjacent frames. Next, in order to normalize the silhouette by person region height, the maximum radius (head point) r_{max} and minimum radius (foot point) r_{min} of the person region in polar coordinate (r, ρ) of the omnidirectional image are found (see Fig. 3(a)). Then, in order to register the horizontal position, median of azimuth angle ρ_{med} of the person region is found (see Fig. 3(a)). Note that radius and azimuth angle are corresponds to vertical and horizontal positions in the panorama image respectively. As a result, head, foot position, and horizontal center in panorama image are represented by Z_{max}, Z_{min}, and ρ_{med} as shown in Fig. 3(b).

Second, silhouette images are scaled so that the height $(Z_{max} - Z_{min})$ in panoramic image can be just 30 pixels, and so that the aspect ratio of each region can be kept. Then, we produce a 20 × 30 pixel-sized image in which the horizontal median ρ_{med} corresponds to the horizontal center of the produced



(a) front-oblique

(b) fronto-parallel

(c) rear-oblique

(d) Definition of observation view

Fig. 4. GSV examples for multiple observation views

image. A GSV is finally constructed by aligning the images on the temporal axis. Figure 4 shows GSV examples for multiple observation views. We can clearly see appearance changes in each view.

3 Multi-view Feature Extraction and Matching

3.1 Frequency-Domain Feature Extraction

The second step in the proposed method is frequency-domain feature extraction from the constructed GSV. First, gait period N_{gait} is detected by maximizing the following normalized autocorrelation $C(N)$

$$C(N) = \frac{\sum_{x,y}\sum_{n=0}^{T(N)} g(x,y,n)g(x,y,n+N)}{\sqrt{\sum_{x,y}\sum_{n=0}^{T(N)} g(x,y,n)^2}\sqrt{\sum_{x,y}\sum_{n=0}^{T(N)} g(x,y,n+N)^2}}, \tag{7}$$

of the GSV $g(x,y,n)$ with the N frame shift for the temporal axis, where N_{total} and $T(N) = N_{total} - N - 1$ is the number of total and overlapped frames in the sequence respectively. The domain of N is set to [25, 45] empirically for natural gait periods. This is because various gait types such as running, brisk walking, and 'ox walking' are not within the scope of this paper.

For the autocorrelation-based period detection, adjacent gait-period sequences need to be similar each other. We assume that a walker's trajectory is smooth to some extent and that appearance changes between adjacent gait-period sequences are small.

Next, the subsequences $\boldsymbol{S}_{n_s}$ is picked up from a complete sequence $\boldsymbol{S}$. Note that the frame range of the subsequence $\boldsymbol{S}_{n_s}$ is $[n_s, n_s + N_{gait} - 1]$. A Discrete Fourier Transformation (DFT) $G_{n_s}(x,y,k)$ for the temporal axis is then applied for the subsequence, and amplitude spectra $A_{n_s}(x,y,k)$ are calculated as

$$G_{n_s}(x,y,k) = \sum_{n=n_s}^{n_s+N_{gait}-1} g(x,y,n)e^{-j\omega_0 kn} \tag{8}$$

$$A_{n_s}(x,y,k) = \frac{1}{N_{gait}}|G_{n_s}(x,y,k)|. \tag{9}$$

where ω_0 is the base angular frequency for the gait period N_{gait}. In this paper, direct-current elements ($k = 0$) (averaged silhouette) and low-frequency elements ($k = 1, 2$) are chosen as experimental gait features. Let $\boldsymbol{a}$ be a feature vector composed of elements of the amplitude spectra $A(x, y, k)$. As a results, the dimension of the feature vector $\boldsymbol{a}$ sums up to $20 \times 30 \times 3 = 1800$.

3.2 Observation View Estimation

In this section, observation view estimation for multi-view feature extraction is addressed. The observation view θ is defined as

$$\theta = (180 - \phi) + \rho \tag{10}$$

where ρ is a azimuth angle, and ϕ is a walking direction (see Fig. 4(d)).

Azimuth angle ρ is simply defined as direction of vector (x, y), where (x, y) is a foot point in the omnidirectional image. Walking direction ϕ is estimated from a trajectory of subject's foot points $F(X, Y)$ on a floor coordinate. Let (R_f, ρ) be polar coordinate on the floor. If the floor plane is regarded as image plane, the distance H_r from mirror focal point O_m to the floor can be seen as focal length to the floor image plane. Then, radius R_f is calculated as follows [11].

$$R_f = \frac{-(b^2 - c^2)H_r r_f}{(b^2 + c^2)f - 2bc\sqrt{r_f^2 + f^2}} \tag{11}$$

Thus, walking trajectory on the floor is obtained as a time series of the above floor points (R_f, ρ).

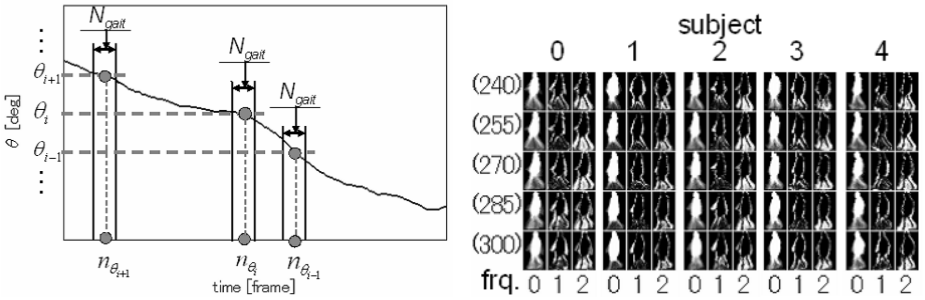
Next, walking direction ϕ is defined as tangential direction of the estimated walking trajectory. Let (X_n, Y_n) and (V_{X_n}, V_{Y_n}) be foot point's position and velocity at nth frame. The velocity is introduced by central difference as follows.

$$V_{X_n} = \frac{X_{n+\Delta n} - X_{n-\Delta n}}{2\Delta n}, \quad V_{Y_n} = \frac{Y_{n+\Delta n} - Y_{n-\Delta n}}{2\Delta n} \tag{12}$$

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_n}, V_{Y_n}).

3.3 Multi-view Feature Extraction

In this section, multi-view feature extraction is introduced based on the estimated observation views. First, multiple basis views $\theta_i (i = 1, 2, \ldots)$ are chosen from observation views. In this time, interval of the basis views is set to 15 deg empirically. 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(a). Concretely speaking, the start frame s in eq. (9) is replaced by $n_s = n_{\theta_i} - N_{gait}/2$.



(a) Overview of multi-view feature extraction

(b) Multi-view features for each subject (every 15 deg)

Fig. 5. Multi-view feature extraction

Results of multi-view feature extraction for multiple subjects are shown in Fig. 5(b). In this figure, each block indicates each subject, and each row and column indicate observation 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.

3.4 Matching Features

A matching measure between two subsequences must first be found if the proposed method is to work. Let S^P and S^G be complete sequences for probe and galley, respectively, and let $S^P_{\theta_i}, S^G_{\theta_i}$ be their subsequences for basis angle θ_i, respectively. Also let $\boldsymbol{a}(\boldsymbol{S}_{\theta_i})$ be feature vector for subsequence $\boldsymbol{S}_{\theta_i}$. The matching measure $d(\boldsymbol{S}_{\theta_i}, \boldsymbol{S}_{\theta_i})$ is simply chosen as the Euclidean distance as $d(S^P_{\theta_i}, S^G_{\theta_i}) = \|\boldsymbol{a}(S^P_{\theta_i}) - \boldsymbol{a}(S^G_{\theta_i})\|$.

Complete sequences have variations in general and may contain outliers. Because a measure candidate $D(\boldsymbol{S_P}, \boldsymbol{S_G})$ can handle this noise, the median value of each subsequence result is used:

$$D(S^P, S^G) = \mathrm{Median}_i\{d(S^P_{\theta_i}, S^G_{\theta_i})\} \tag{13}$$

4 Experiment

4.1 Datasets

A total of 60 gait sequences from 15 subjects were used for the experiments. Each sequence consisted of approximately 10 steps of a straight walk in front of the omnidirectional camera, and it included 5 basis views: 240, 255, 270, 285, and 300 deg. The camera was Sony Inc. DCR-VX2000, and images were captured by 720 × 480 pixel size at 30 fps. The hyperboloidal mirror and camera parameters were $a = 13.722$C$b = 11.708$C$c = 18.038$C$f = 427.944$ (unit: mm). The dataset was taken for two days, that is, there were two sequences per day for each subject. A test set is composed of one gallery sequence of a day and two probe sequences of the other day, therefore totally four combinations of test sets were generated.

4.2 Results

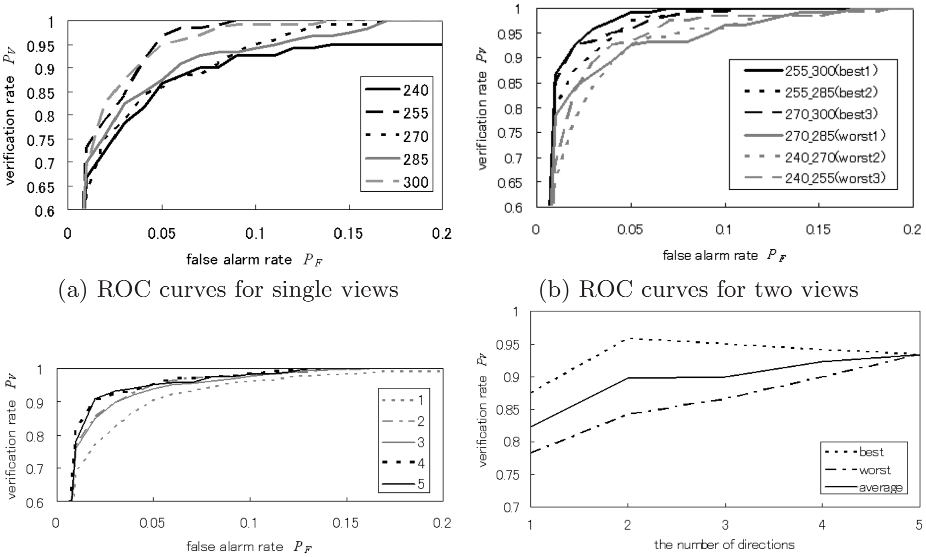
The gait identification experiments were done for the above four combinations of datasets and average performance was evaluated by Receiver Operating Characteristics (ROC) curves [12]. The ROC curves shows relation between verification rate P_V and false alarm rate P_F when the receiver changes the acceptance thresholds. The ROC curves tilting toward left top corner in the graph indicates high performance because it means high verification rate at low false alarm rate. In addition, the effect of the number of observation views and combinations of views on performance are analyzed to validate the effectiveness of multi-view observations.

First, ROC curves for each single view are illustrated in Fig. 6(a). The figure shows that performance varies greatly among basis views, and that it is difficult to gain enough performance when an arbitrary single-view feature is used for matching.

Next, ROC curves for two-view combinations are illustrated in Fig. 6(b). Here, the best and the worst three combinations are shown. Note that the performance order is judged by Equal Error Rate (EER), that is, error rate when false alarm rate P_F becomes equal to false rejection rate $(1 - P_V)$. Focusing on the worst cases, view differences are small (within 15 deg except for the worst 2). On the other hand, focusing on the best cases, view differences are relatively large (more than 30 deg). As a result, it is clear that the combination with large view difference is effective for identification.

Moreover, ROC curves for each number of observation views are shown in Fig. 6(c). In verification rates in this graph are averaged over all the combinations for each number of observation views. As a result, we can see that the performance becomes better as the number of observation views increase.

Finally, verification rates at 3% false alarm rate are picked up for each number of observation views. Figure 6(d) shows of the best, the worst, and the averaged performance over all the combinations. As for the best combinations, the verification rate for two observation views reaches the highest performance. Thus a small number of observation views are enough when the combination can be specified. As for the worst combination, the verification rates make a steady progress as the number of observation views increase. In case of the worst, because combinations are usually composed of adjacent views as known from two views combination case, the increase of the number of observation views directly leads to observation views variation. In summary, it is validated that observation view variations greatly contribute to performance improvement.



(a) ROC curves for single views

(b) ROC curves for two views

(c) ROC curves for each number of views

(d) Verification rate at 3% false alarm for each number of views

Fig. 6. Experimental results

5 Conclusion

This paper describes a method of gait identification based on multi-view gait images using an omnidirectional camera.

The omnidirectional silhouette images first transformed into panoramic ones and a spatio-temporal Gait Silhouette Volume (GSV) is obtained. Next, frequency-domain features are extracted by Fourier analysis. Because the omnidirectional camera makes it possible to observe a person from various views, multi-view features can be extracted from the GSVs composed of multi-view images. In an identification phase, distance between a probe and a gallery feature of the same view is calculated, and then these for all views are integrated for matching. The effect of observation view variation on gait identification performance was analyzed through experiments including 15 subjects from 5 views. As a result, average performance increases from 82% (single view) to 93% (5 views), and it was clear that observation view variation contributes to gait identification performance.

In this paper, basis views are chosen only from views common for a gallery and a probe. It is possible to use other-view features interpolated by View Transformation Model (VTM) [8] for better performance in a future work. Moreover, subjects in this experiment walked within 5m from the omnidirectional camera, and thus relatively high-resolution silhouettes (approximately 60 pixel height) are obtained. Therefore, effects of distance from the camera or silhouette resolution on identification performance should be analyzed. That also leads to analysis of the optimal alignment of the omnidirectional camera to capture multi-view gait images effectively considering both silhouette resolution and observation view variation.

References

1. Urtasun, R., Fua, P.: 3d tracking for gait characterization and recognition. In: Proc. of the 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 17–22. IEEE Computer Society Press, Los Alamitos (2004)
2. Yam, C., Nixon, M., Carter, J.: Automated person recognition by walking and running via model-based approaches. Pattern Recognition 37(5), 1057–1072 (2004)
3. 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(2), 162–177 (2005)
4. Han, J., Bhanu, B.: Individual recognition using gait energy image. Trans. on Pattern Analysis and Machine Intelligence 28(2), 316–322 (2006)
5. 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, vol. 1, pp. 807–816 (2006)
6. 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, pp. 143–150. IEEE Computer Society Press, Los Alamitos (2003)
7. Shakhnarovich, G., Lee, L., Darrell, T.: Integrated face and gait recognition from multiple views. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 439–446 (2001)
8. 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, Graz, Austria, vol. 3, pp. 151–163 (2006)
9. 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)
10. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Adaptation to walking direction changes for gait identification. In: Proc. of the 18th Int. Conf. on Pattern Recognition, Hong Kong, China, vol. 2, pp. 96–99 (2006)
11. Yamazawa, K., Yagi, Y., Yachida, M.: Hyperomni vision: Visual navigation with an omnidirectional image sensor. Systems and Computers in Japan 28(4), 36–47 (1997)
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(10), 1090–1104 (2000)