Matching Gait Image Sequences in the Frequency Domain for Tracking People at a Distance

Ryusuke Sagawa, Yasushi Makihara, Tomio Echigo, and Yasushi Yagi

Institute of Scientific and Industrial Research, Osaka University,
8-1 Mihogaoka, Ibaraki-shi, Osaka, 567-0047, JAPAN
{sagawa,echigo,yagi}@am.sanken.osaka-u.ac.jp

Abstract. This paper describes a new method to track people walking by matching their gait image sequences in the frequency domain. When a person walks at a distance from a camera, that person often appears and disappears due to being occluded by other people and/or objects, or by going out of the field of view. Therefore, it is important to track the person by taking correspondence of the image sequences between before and after the disappearance. In the case of tracking, the computational time is more crucial factor than that in the case of identification. We create a three-dimensional volume by piling up an image sequence of human walking. After using Fourier analysis to extract the frequency characteristics of the volume, our method computes the similarity of two volumes. We propose a method to compute their correlation of the amplitude of the principal frequencies to improve the cost of comparison. Finally, we experimentally test our method and validate that the amplitude of principal frequencies and spatial information are important to discriminate gait image sequences.

1 Introduction

When a surveillance system using cameras tracks people who walk at a distance from the cameras, the subjects are often occluded by other people and/or objects. Therefore, it is necessary to make correspondences of tracked people between before and after their occlusion. For example, if a multiple-camera system tracks people as shown in Figure 1, making correspondences of objects is necessary between before and after the occluded area. Since the images of people at a distance from a camera are small, it is difficult to recognize them by their facial appearance. Though color and shape are considered cues for matching, this paper focuses on the gaits of people, which are also important features. When walking, people move their torso, arms, and legs in a unique way. Hence the rhythm of a gait is different among individuals. Since gait can be observed at a distance, gait matching has advantages for a tracking system.

The issue of matching gait image sequence for tracking is similar to the identification problem of gait image sequence. However, the differences from identification are as follows:

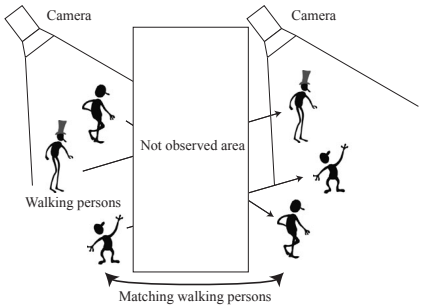


Fig. 1. Making correspondences of pedestrians with a multiple-camera system.

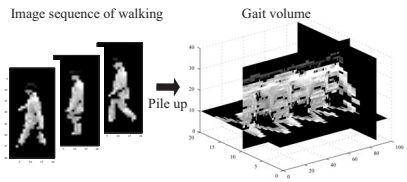


Fig. 2. A gait volume is created by piling up the image sequences of walking.

- The features should be extracted only from an image sequence.
- The computational time is very important.

In the case of tracking problem, the features of walking should be extracted only from an image sequence while multiple image sequences can be used from a gait database in the case of identification. Moreover, the computational time is more crucial factor than that in the case of identification.

Several approaches have been proposed for identification of a person from their gait. They can mostly be classified into two classes, model- or appearance-based approaches. Model-based approaches extract the motion of the human body by fitting their models to input images. Yam et al. [1] and Cunado et al. [2] extracted leg motions and found their gait signature by Fourier analysis. Urtasun and Fua [3] used a 3D temporal motion model to increase the robustness for a changing view direction. Bobick and Johnson [4] extracted activity-specific static body parameters instead of directly analyzing gait motion. Lee and Grimson [5] analysed the frequency fo 7 parts which are extracted from the silhouette of human walking motion.

Appearance-based approaches directly extract parameters from images without assuming a model of a human body and its motion. Niyogi and Adelson [6] used 3D spatio-temporal (XYT) data by piling up images and extracted gait motion by fitting a 'snake'. Murase and Sakai [7] proposed a template matching method in the parametric eigenspace that is projected from images. Little and Boyd [8] recognized individuals by frequencies and phases computed by extracting optical flows. Liu and Picard [9] used a spatio-temporal volume and detected the periodicity in a motion by 1D Fourier analysis for each pixel of the image. BenAbdelkader et al. [10] used self-similarity plots, in which each pixel had correlations with the frames of an image sequence. Liu et al. [11] used a frieze pattern to represent gait motion; a pattern created by summing up the white pixels of a binarized image of a gait along the rows and columns of an image. Sarker et al. [12] proposed a baseline algorithm of gait recognition, which computes the similarity of gait sequences by spatial-temporal correlation. Han and Bhanu [13] proposed a representation of gait image sequence, called a gait energy image, which is computed by taking average of a silhouette image sequence.

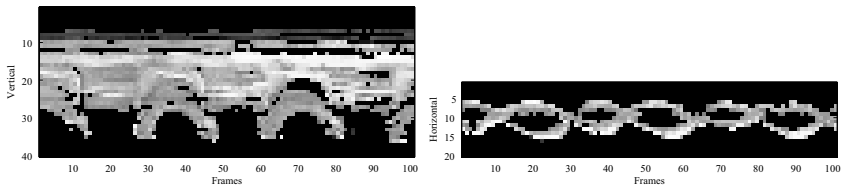
The previous model-based approaches to determine the frequency of a gait only considered the frequency of some parts of the body by extracting them from image sequence. However, we think that individuality is represented by a mixture of frequencies of whole body. Therefore, we attempted to extract the distribution of the various frequencies included in the motion of every part of a walking body. Our proposed approach creates 3D spatio-temporal volume from an image sequence, which is similar to Niyogi's [6] and Liu's [9] methods. Spatio-temporal volume data, here called *gait volume*, contain information not only of spatial individuality such as features of the torso and face, but also the movement of the body with its unique rhythm. By extracting the frequency characteristics of the volume by computing 1-D Fourier transform along a time axis, our method computes the similarity of two volumes. Though we analyze it by 3D Fourier transform in our previous paper [14], the spatial information is omitted. Thus, a new method utilizes the spatial information for matching. We propose a method to compute their correlation of the amplitude. Since it is not necessary for this method to align frames for matching two sequences, this method has advantage with respect to the computational cost.

In the following sections, we describe details of our method. In Section 2, we explain how a gait volume is created. In Section 3, we describe a method to extract frequency information by 1-D Fourier transform. Next, in Section 4, our method of matching frequency information is described. We experimentally test the proposed method in Section 5 and summarize our contribution in Section 6.

2 Creating a Spatio-Temporal Volume

A gait volume is created by piling up image sequences of a person walking as shown in Figure 2. The process consists of two steps: background subtraction and image alignment. The human region is extracted by subtracting background from input images [15]. Moving regions are extracted as the human region by subtracting the stationary background images from input ones by pixels. After extracting the human region, the principle axis of the body is calculated as a horizontal position in the human region in an image if it is assumed that a person walks in a fronto-parallel plane. A sequence of the extracted human region is then temporally aligned by shifting the extracted human region. Without specifying each of the parts of the body, we can create a gait volume that contains the continuous changes of appearance while walking. These changes exactly reflect the gait rhythm.

A gait volume includes both spatial and temporal information. Sliced planes of the volume data express changes of textures in the subject's walking, and also represent the rhythm in a person's gait. Figure 3 shows examples of vertical and horizontal slices of a gait volume. From the slices, it is possible to acquire information about how a person moves his/her body while walking. Figure 3(a) is a vertical slice at the central column. There are vertical waves around the shoulder, arms and waist. Figure 3(b) is a horizontal slice at the row of the knee



(a) (b)

Fig. 3. (a) Vertical and (b) horizontal slices of a gait volume.

level. There are horizontal waves of legs and the difference in the motions of the right and left legs can be observed.

3 Frequency Analysis of Gait Volume

In this section, we extract the frequency characteristics of a gait volume by Fourier transform. Our method consists of three steps:

1. Compute Fourier transform $G(x, y, k)$ of a gait volume $g(x, y, n)$ along the time axis.
2. Extract the principal frequencies of a gait volume.
3. Remove spectra from $G(x, y, k)$ other than the principal frequencies.

First, we compute 1-D discrete Fourier transform for each pixel of images along the frame axis:

$$G(x, y, k) = \sum_{n=0}^{N-1} g(x, y, n) \exp(-\frac{2\pi ikn}{N}), \tag{1}$$

where $g(x, y, n)$ is the intensity of a pixel (x, y) at n-th frame and N is the number of frames. Figure 4 shows an example of the change of the intensity of a pixel in a gait volume. After Fourier transform, the amplitude of the pixel becomes as shown in Figure 5. The frequency f corresponding to k is computed as $f = \frac{k}{N\Delta t}$, where Δt is the sampling interval of images.

Second, since the amplitudes of the most of frequencies are small while the dimension of $G(x, y, k)$ is very high, we extract the principal frequencies of a gait, which have large amplitudes, to reduce the data size and improve the computational cost. We compute the sum of the amplitude of $G(x, y, k)$ for each frequency:

$$\hat{G}(k) = \sum_{x,y} |G(x, y, k)|. \tag{2}$$

Since $G(x, y, k)$ is a complex value, $|G(x, y, k)| = \sqrt{a^2 + b^2}$, where $G(x, y, k) = a + bi$. Then, we find the principal frequency of $G(x, y, k)$ as the frequency k that satisfies $\hat{G}(k-1) < \hat{G}(k)$ and $\hat{G}(k+1) < \hat{G}(k)$. Since the higher frequencies is not important, we choose some lower frequencies from them. The DC component

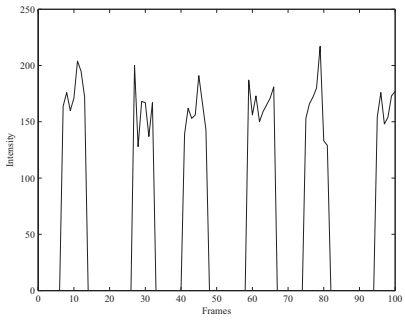


Fig. 4. Change of the intensity of a pixel in a gait volume.

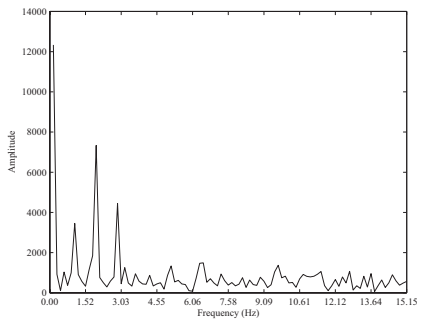


Fig. 5. Amplitude of a pixel after Fourier transform.

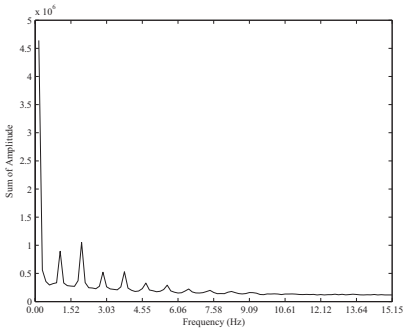


Fig. 6. Sum of the amplitude of $G(x, y, k)$.

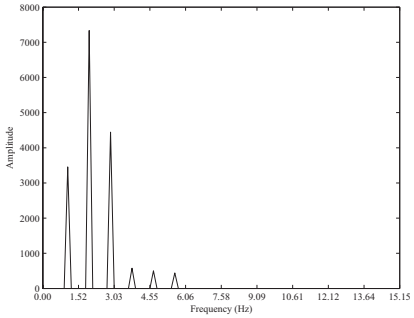


Fig. 7. Example of amplitudes of principal frequencies.

is ignored for this computation because it does not represent a repetitive motion of walking. Figure 6 shows an example of $\hat{G}(k)$. There are some peaks and we extract their frequencies as the principal frequencies.

In the third step, we remove spectra included in $G(x, y, k)$ other than the principal frequencies and obtain a new volume $G'(x, y, k)$. We preserve the several lowest principal frequencies and remove all other frequencies. Figure 7 shows the amplitude of $G'(x, y, k)$ after removing spectra other than the principal frequencies from Figure 5.

Figure 8 shows the reconstructed results of the following inverse Fourier transform:

$$g'(x, y, n) = \frac{1}{N} \sum_{k=0}^{N-1} G'(x, y, k) \exp(\frac{2\pi i k n}{N}). \tag{3}$$

Figure 8(a) is an original image in a gait volume and Figure 8(b) is the reconstructed image from $G'(x, y, k)$. Figure 8(c) shows the amplitudes of three principal frequencies of $G'(x, y, k)$. Figure 8(d) is the horizontal slice of $g'(x, y, n)$, which corresponds to Figure 3(b).

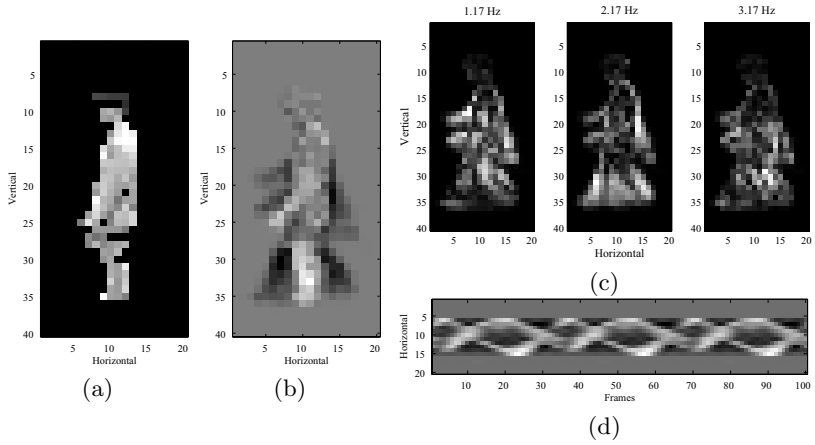


Fig. 8. (a) Original image, (b) reconstructed image by inverse Fourier transform, (c) Amplitudes of three principal frequencies, and (d) the horizontal slice of the reconstructed gait volume which corresponds to Figure 3(b).

4 Matching Gait Volumes

To compare two different gait volumes, we propose a method for computing their correlation. We use the correlation of the amplitude after Fourier transform. In this section, we assume that the images of walking people are aligned along the horizontal and vertical axis of gait volumes.

Now, $G_1(x, y, k)$ and $G_2(x, y, k)$ are the volumes after Fourier transform. If G_1 is a reference volume, we remove spectra other than the principal frequencies of G_1 from both G_1 and G_2, and obtain $G'_1(x, y, k)$ and $G'_2(x, y, k)$. Namely, G'_1 and G'_2 only have the principal frequencies of G_1. We compare the amplitude of G'_1 and G'_2 by the following criterion:

$$C(|G'_1|, |G'_2|) = \frac{\sum_{x,y,k} |G'_1(x, y, k)||G'_2(x, y, k)|}{\sqrt{(\sum_{x,y,k} |G'_1(x, y, k)|^2)(\sum_{x,y,k} |G'_2(x, y, k))|^2}}. \quad (4)$$

This is the normalized correlation without shifting by mean value. Thus, if two volumes are same, the result is 1, and it goes down to -1 if they have negative correlation.

Removing spectra other than the principal frequencies is equal to reducing the dimension of the component in gait volumes. Therefore, the cost of computing (4) is much smaller than that of computing the correlation of the original volumes by $C(|G_1|, |G_2|)$.

5 Experiments

We first tested our method with image sequences in which people walk in a fronto-parallel plane to the camera. We used 50 sequences of 9 persons, i.e., 4-7

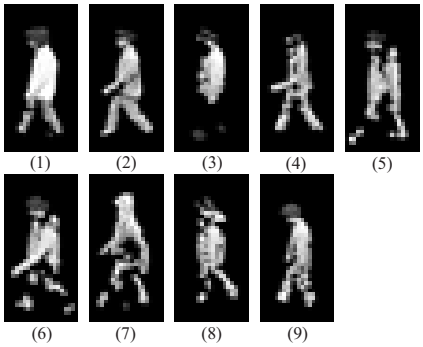


Fig. 9. Subject images.

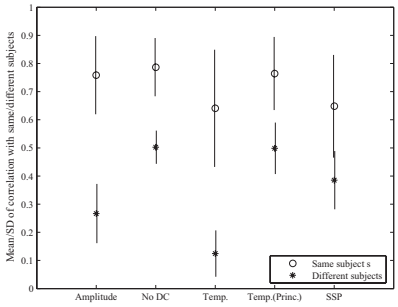


Fig. 10. Comparison of five methods: the means of comparing the same subjects are indicated by ∘, while those of comparing different subjects are indicated by *. Vertical lines show $\pm\sigma$, which is the standard deviation.

sequences for each person. Each sequence consists of 200 frames, captured at 33Hz, and include 10-12 walking steps. The size of the images is 40×20. Figure 9 shows images of the subjects after background subtraction.

We compute the correlation of all pairs of sequences. To evaluate the efficiency of our method, we compared the following five methods:

- Amplitude: the correlation of $|G'(x, y, k)|$ proposed in Section 4.
- No DC: the correlation of $|G(x, y, k)|$ by (4) simply after removing the DC component.
- Temp.: normalized correlation in the spatio-temporal domain of $g(x, y, n)$.
- Temp.(Princ.): normalized correlation in the spatio-temporal domain of $g'(x, y, n)$, which is obtained by inverse Fourier transform of $G'(x, y, k)$.
- SSP: normalized correlation of the self-similarity plots proposed in [10].

In this experiment, we used the three lowest principal frequencies for matching. Thus, the data size is reduced to 3% of the original gait volume. Since a self-similarity plot is a matrix of the correlation of two images in a image sequence, the spatial information is lost in this representation. In the SSP method, we compute the normalized correlation of the self-similarity plots generated from gait sequences. Though the method of comparing the self-similarity plots is different from the one proposed in [10], we compare the effectiveness as a cue for identification by normalized correlation. In the Temp., Temp.(Princ.) and SSP methods, we search for the best match by an exhaustive brute-force search with circular shift in the spatio-temporal domain.

We apply the five above methods to all pairs of gait sequences. Figure 10 shows the mean and standard deviations of the correlations of each method when they are compared to the sequences of the same and different subjects. The means of comparing the same subjects are indicated by ∘ while those of

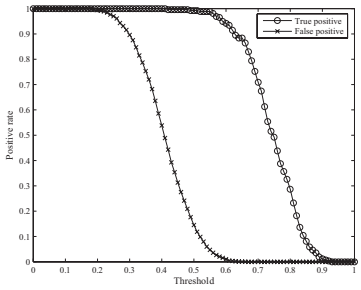


Fig. 11. Rates of positive samples after thresholding by correlation values.

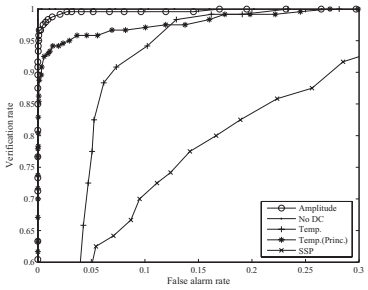


Fig. 12. ROC curves for five methods.

comparing different subjects are indicated by *. The vertical lines shows $\pm\sigma$, which is the standard deviation.

If the difference between $\circ$ and * is large, the method is effective to distinguish the gaits of different persons. In the Amplitude method, the difference is sufficiently large compared to their standard deviations. Therefore, the method can be seen as effective for comparing gait image sequences. Since the difference of the SSP method is small while the standard deviation is large, the robustness for matching is worse than others. Therefore, the spatial information in an image is important even if the frequency information is used for matching.

Figure 11 shows the rate of positive samples after thresholding by correlation values for the Amplitude method. The true positive is the result of matching the same subjects, and the false positive is that of matching different subjects. When the rate of true positive is 0.95, that of false positive is less than 0.01. Thus, the Amplitude method can discriminate subjects by thresholding the correlation values.

We evaluate the rates of positive samples for five methods, and create the receiver operating characteristic (ROC) curves as shown in Figure 12. It depicts the relationship of true and false positive rates. The No DC method has no error in this experiment. Therefore, it is shown that the frequency information is a powerful feature for matching. Though the data size for the Amplitude method is quite smaller than the No DC method, it has few errors. Hence, it is effective to use the principal frequencies for matching gait volumes. As for the Temp. and Temp.(Princ.) method, the result is worse than the No DC and Amplitude method. It shows that the template matching in the temporal domain is not suitable for matching gait volumes.

Table 1 shows the computational time for comparing a pair of gait sequences by these five methods. We used a PC with Pentium4 3.2GHz processor and coded the algorithms by MATLAB. The time of the Amplitude method is 16% of the No DC method. Thus, the computational cost is reduced by removing the minor component in $G(x, y, k)$. Since the temporal alignment is necessary for the other methods, their computational cost is higher than that of the Amplitude method.

Table 1. Times for comparing a pair of gait sequences in seconds.

Amplitude	No DC	Temp.	Temp.(Princ.)	SSP
0.015	0.093	4.0	4.0	2.5

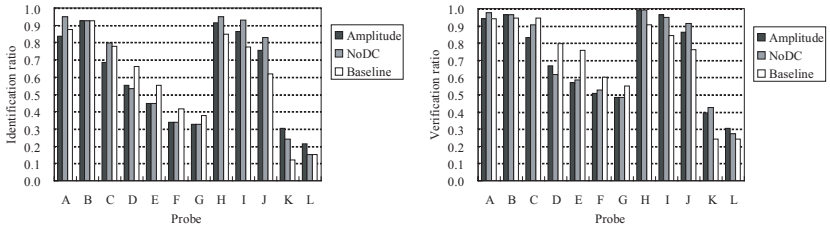


Fig. 13. Comparison of Amplitude, No DC and Baseline methods for USF database: (a) the identification rate at rank 5, (b) the verification rate at a false alarm rate of 10%. The differences between gallery and probe sequence: (A) view, (B) shoe, (C) shoe, view, (D) surface, (E) surface, shoe, (F) surface, view, (G) surface, shoe, view, (H) briefcase, (I) shoe, briefcase, (J) view, briefcase, (K) time, shoe, clothing, (L) surface, time, shoe, clothing.

Second, we tested our method by using a database of gait image sequence from University of South Florida; for details of the database, refer to [12]. The database consists of gallery (watch-list) and probe (input data) image sequences, which are compared in the experiment. We used the silhouettes which are already extracted by their algorithm in this experiment. The number of gallery sequences is 121. The size of images we use is normalized to 88×128 pixels, and the number of frames for matching is 128. We compared three methods, Amplitude, No DC and Baseline [12]. Figure 13 shows identification and verification rates for each probe. The difference between Amplitude and No DC methods is small while the cost of Amplitude method is much smaller than No DC method. Moreover, the costs of these methods are much smaller than Baseline method because aligning frames is necessary for Baseline method. Though the performance of Amplitude and No DC methods becomes worse than Baseline method for (B)-(G) probes, which have difference about surface, it is considered to be due to background subtraction. On the other hand, our method has advantage for (H)-(J) probes, which have difference about one's belongings. It is considered that the frequency information is not affected by carrying briefcase.

6 Summary

We proposed a new method to compare gait image sequences. The characteristics of the gait are extracted from a gait volume using Fourier transform. We use the principal frequencies in the frequency domain for matching gait volumes. Thus, the data size and computational cost become quite smaller than the original

gait volume. It works better than matching in the temporal domain, and the computational cost is small because the temporal alignment is not necessary. This advantage is suitable for tracking problem. Moreover, it is shown that the spatial information is also important to discriminate gait image sequences. For future work, we analyze the effect of other factors, for example, a viewing direction and clothes.

References

1. Yam, C., Nixon, M., Carter, J.: Automated person recognition by walking and running via model-based approaches. Pattern Recognition **37** (2004) 1057–1072
2. Cunado, D., Nixon, M., Carter, J.: Automatic extraction and description of human gait models for recognition purposes. Computer Vision and Image Understanding **90** (2003) 1–41
3. Urtasun, R., Fua, P.: 3d tracking for gait characterization and recognition. In: Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition. (2004) 17–22
4. Bobick, A., Johnson, A.: Gait recognition using static activity-specific parameters. In: Proc. Computer Vision and Pattern Recognition. (2001)
5. Lee, L., Grimson, W.: Gait analysis for recognition and classification. In: Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition. (2002) 155–162
6. Niyogi, S., Adelson, E.: Analyzing and recognizing walking figures in xyt. In: Proc. Computer Vision and Pattern Recognition. (1994) 469–474
7. Murase, H., Sakai, R.: Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern Recognition Letters **17** (1996) 155–162
8. Little, J., Boyd, J.: Recognizing people by their gait: The shape of motion. Videre **1** (1998)
9. Liu, F., Picard, R.: Finding periodicity in space and time. In: Proc. the Sixth International Conference on Computer Vision. (1998) 376–383
10. BenAbdelkader, C., Culter, R., Nanda, H., Davis, L.: Eigengait: Motion-based recognition people using image self-similarity. In: Proc. AVBPA. (2001)
11. Liu, Y., Collins, R., Tsin, Y.: Gait sequence analysis using frieze patterns. In: Proc. the 7th European Conference on Computer Vision. Volume 2. (2002) 657–671
12. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: Data sets, performance, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 162–177
13. Han, J., Bhanu, B.: Statistical feature fusion for gait-based human recognition. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2004) 842–847
14. Ohara, Y., Sagawa, R., Echigo, T., Yagi, Y.: Gait volume: Spatio-temporal analysis of walking. In: Proc. The fifth Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras (OMNIVIS2004), Prague, Czech (2004)
15. Mituyosi, T., Yagi, Y., Yachida, M.: Real-time human feature acquisition and human tracking by omnidirectional image sensor. In: Proc. IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems. (2003) 258–263