Person identification from spatio-temporal 3D gait

Yumi Iwashita Ryosuke Baba Koichi Ogawara
Ryo Kurazume
Information Science and Electrical Engineering
Kyushu University, Japan
yumi@ait.kyushu-u.ac.jp

Abstract

This paper presents a spatio-temporal 3D gait database and a view independent person identification method from gait. In case that a target changes one's walking direction compared with that in a database, the correct classification rate is reduced due to the appearance change. To deal with this problem, several methods based on a view transformation model, which converts walking images from one direction to virtual images from different viewpoints, have been proposed. However, the converted image may not coincide the real one, since the target is not included in the training dataset to obtain the transformation model. So we propose a view independent person identification method which creates a database with virtual images synthesized directly from the target's 3D model. In the proposed method, firstly we built a spatio-temporal 3D gait database using multiple cameras, which consists of sequential 3D models of multiple walking people. Then virtual images from multiple arbitrary viewpoints are synthesized from 3D models, and affine moment invariants are derived from virtual images as gait features. In the identification phase, images of a target who walks in an arbitrary direction are taken from one camera, and then gait features are calculated. Finally the person is identified and one's walking direction is estimated. Experiments using the spatio-temporal 3D gait database show the effectiveness of the proposed method.

1 Introduction

Person recognition systems have been used for a wide variety applications, such as surveillance applications for intelligence and security operations. Several person recognition systems using biometrics, such as fingerprints and iris, are available in the market. However, these systems need special equipment and require interaction with or cooperation of the subject. On the other hand, gait recognition has the advantage of being unobtrusive because body-invasive sensing is not needed to capture gait information. Moreover, gait recognition has the extra advantage that it may be performed from a distance.

Gait recognition approaches generally fall into two main categories: (1) model-based structural analysis, and (2) appearance-based analysis. Structural model-based approaches include parameterization of gait dynamics, such as stride length, cadence, and joint angles [2] [3]. Traditionally, these approaches have not reported high performances on common databases, partly due to their need for 3D calibration information and self-occlusion caused by legs and arms crossing.

Appearance-based analysis uses measurements of spatio-temporal features of silhouettes [1] [6] [8], and shape and its variations are characterized by feature extraction methods, such as Fourier transforms [1] [10], and affine moment invariants [6]. Appearance-based approaches have been used with good results on human identification. However, since in general common gait databases are created using one camera and conventional methods assume walking direction in the database is the same with that in test data sets, the correct classification rate is reduced in case target's walking direction is different from that in the database as shown in Fig. 1.

To overcome this problem, Kale et al. introduced a view transformation method based on perspective projection of the sagittal plane [7]. This method can deal with the change of azimuth angle, but cannot do the change of elevation angle. Moreover, Makihara et al. [10] introduced a view transformation model to synthesize virtual viewpoint images from captured images. In this method, the view transformation model is obtained from a training set of multiple people from multiple view directions. Since the target is not included in the training set, so the synthesized image may not coincide to the real one.

In this paper, we propose a view independent person identification method which creates a database with virtual images synthesized directly from the target's 3D model. In the proposed method, firstly we build a spatio-temporal 3D gait database which consists of sequential 3D models

978-0-7695-4175-4/10 $26.00 © 2010 IEEE
DOI 10.1109/EST.2010.19

of multiple people reconstructed with 16 cameras. Virtual images from multiple arbitrary viewpoints are synthesized from 3D models, and affine moment invariants are derived from virtual images at each viewpoint as gait features. In the identification phase, images of a target who walks in an arbitrary direction are taken from one camera, and then gait features are calculated. Finally by comparing the gait features with those in the database, the person is identified and one's walking direction is estimated.

This paper is organized as follows. Section 2 introduces the spatio-temporal 3D gait database, and Section 3 describes the details of the proposed view independent person identification method. Section 4 describes experiments performed using the spatio-temporal 3D gait database. Conclusions are presented in section 5.

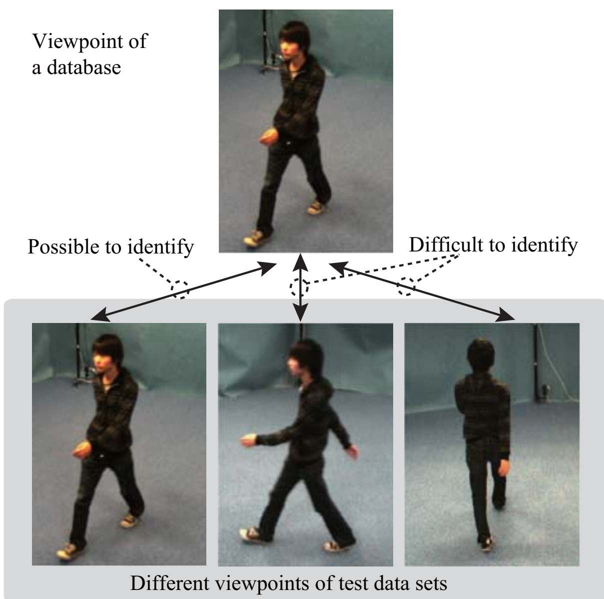


Figure 1. Influence to the correct classification rate due to walking direction changes

2 Spatio-temporal 3D gait database

In this section we introduce the spatio-temporal 3D gait database. To create the database, firstly we build a cylinder-like studio whose radius and height are 3.5 m and 2.6 m, respectively (Fig.2). Eight poles were placed nearly evenly on the circumference of the studio, and we placed 2 cameras at each pole, totally 16 cameras (PointGrey, Dragonfly2). The image resolution is 1032×776, and every camera captured images every 66 msec. Here, internal and external camera parameters were estimated in advance. Ten people with 4 sequences for every person were recorded, and they walked straight during the recording.

To reconstruct the 3D models of the target from walking images, firstly silhouettes from walking images are extracted by a background subtraction. Then the 3D models are reconstructed by the volumetric intersection technique [9]. Figure 3(a) and (c) show examples of reconstructed 3D models of two targets, and Fig. 3(b) and (d) show examples of sequential 3D models of one gait cycle. Here, one gait cycle is a fundamental unit to describe the gait during ambulation, which occurs from the time when the heel of one foot strikes the ground to the time at which the same foot contacts the ground again.

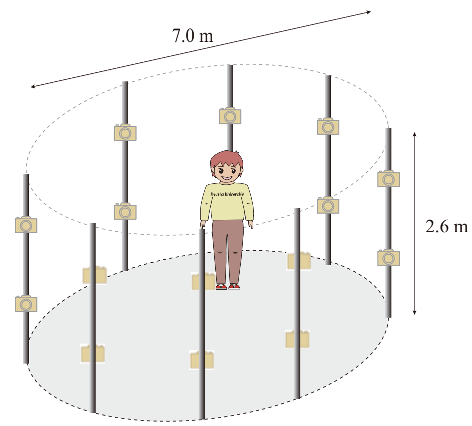


Figure 2. The studio

3 View independent person identification method

In this section we describe the details of the view independent person identification method. To summarize, the main steps of processing are as follows.

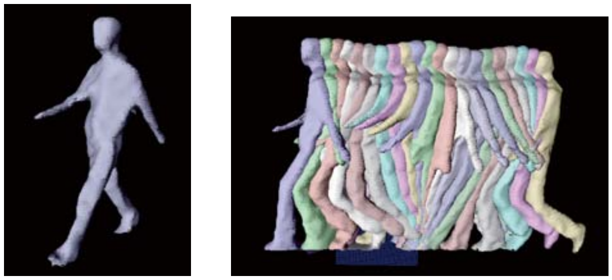
Step 1 Virtual images from an arbitrary viewpoints are synthesized from 3D models in the database as shown in Fig. 4.

Step 2 Affine moment invariants are calculated as gait features from synthesized walking images of one gait cycle.

Step 3 Repeat the Step 1 and Step 2 at all arbitrary viewpoints, and then create the database.

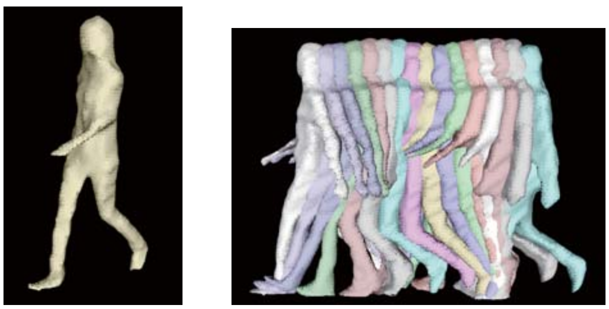
Step 4 In the identification phase, images of a target who walks in an arbitrary direction are captured with one camera, and the gait features of the target are calculated by the same procedure of Step 2.

Step 5 By comparing the gait features of Step 4 with those in the database, the target is identified and one's walking direction is estimated.



(a) (b)

Subject A



(c) (d)

Subject B

Figure 3. (a) and (c): examples of 3D models, (b) and (d): examples of sequential 3D models.

The following offers more details.

3.1 Synthesis of virtual images from arbitrary viewpoints

Virtual images are synthesized by projecting all 3D models in the database onto 2D images of multiple arbitrary viewpoints for each 3D model. Figure 5 shows examples of virtual images, which are synthesized by projecting a 3D model to 2D images of viewpoints at every 5 degree from 0 to 180.

3.2 Feature extraction from virtual images

From synthesized virtual images affine moment invariants [6] are calculated as gait features. A Fourier transform-based method enabled to identify people with high correct classification rate, but the dimension number of features rises in parallel with the increase of the image resolution. On the other hand, the person identification method using affine moment invariants enables to identify people

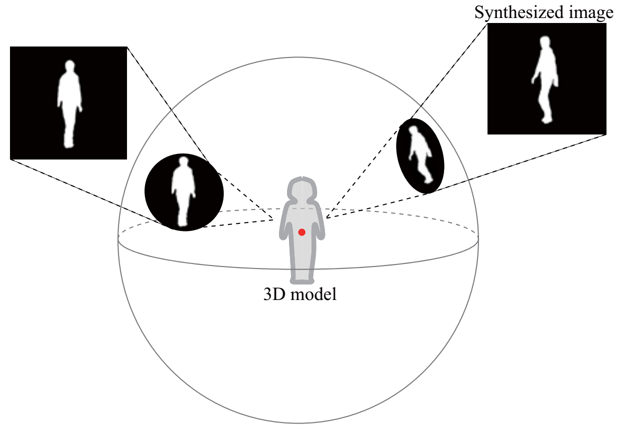


Figure 4. Synthesize a virtual images from an arbitrary viewpoint.

with high correct classification using one twenty-seventh dimension number of features of the Fourier transform-based method. The details of the feature extraction using affine moment invariants are provided as follows.

Firstly, from virtual images of one gait cycle at an arbitrary viewpoint, the average image is defined as follows:

$$I_{average}(x,y) = \frac{1}{T}\sum_{t=1}^{T} I(x,y,t), \qquad (1)$$

where T is the number of frames in one gait cycle and $I(x,y,t)$ represents the intensity at time t. Moreover, the shadow area is scaled to a uniform height. Figure 6 shows an example of average images.

Next, affine moment invariants are calculated from average images. Affine moment invariants are moment-based descriptors, which are invariant under a general affine transform. The moments describe shape properties of an object as it appears. For an image the centralized moment of order $(p+q)$ of an object O is given by

$$\mu_{pq} = \sum\sum\nolimits_{(x,y)\in O} (x-x_g)^p (y-y_g)^q I(x,y). \qquad (2)$$

Here, x_g and y_g are the center of gravity of the object. Six affine moment invariants are listed below[4][5].

$$\begin{aligned}
I_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2) \\
I_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 \\
&\quad +4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
I_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
&\quad +\mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2))
\end{aligned}$$

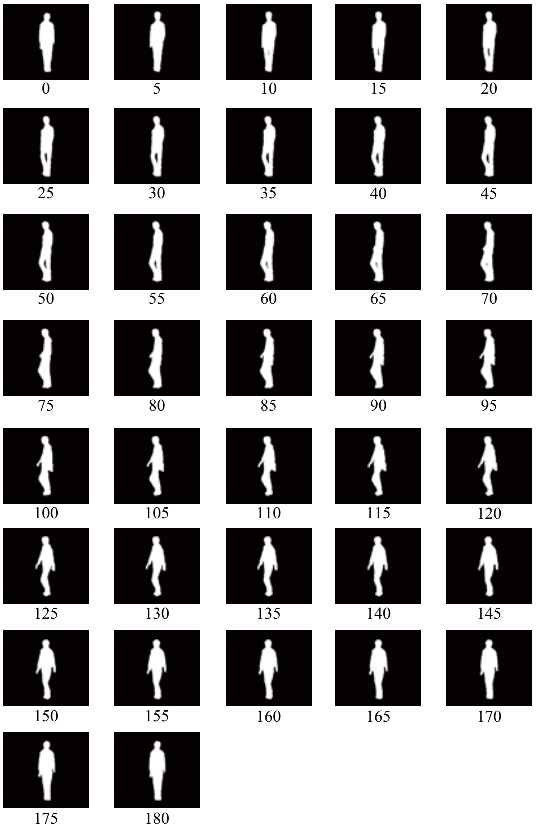


Figure 5. Virtual images synthesized at arbitrary viewpoints

$$
\begin{aligned}
I_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} \\
&\quad +9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} \\
&\quad +6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\
&\quad -8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad +12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2) \\
I_5 &= \frac{1}{\mu_{00}^{6}}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) \\
I_6 &= \frac{1}{\mu_{00}^{9}}(\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{40}\mu_{13}^2 \\
&\quad -\mu_{04}\mu_{31}^2 - \mu_{22}^3)
\end{aligned} \tag{3}
$$

Moreover, since different people move their arms differently when they walk, we divide average images into upper and lower regions at the center of the y-axis to extract features of legs and arms separately.

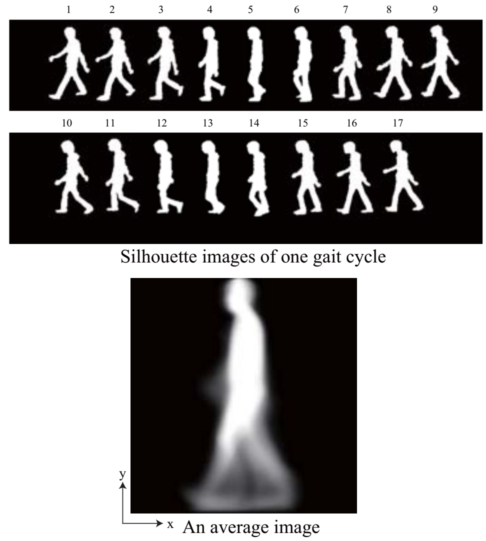


Figure 6. An example average images

3.3 Person identification

To identify a target, firstly one's images who walks in an arbitrary direction are captured with one camera, and then extract silhouettes from images by a background subtraction. Then, the average image is created from captured images of one gait cycle, and calculate affine moment invariants as gait features.

Next, affine moment invariants at each arbitrary viewpoint in the training data sets are whitened, and build the database with whitened affine moment invariants. In the identification phase, affine moment invariants of the target are whitened, and then the target is identified and one's walking direction is estimated by a classifier. In experiments, we use the k-nearest neighbor (knn) as the classifier.

3.4 Characteristics of the proposed method

The characteristics of the proposed method is as follows. The proposed method can identify the target robustly against walking direction changes by creating the database with virtual images synthesized from 3D models. Moreover, people can be identified with high correct classification using small dimension number of features by using affine moment invariants.

4 Experiments

In this section, we describe the experiments. In our experiments, we applied the proposed method to the spatio-temporal 3D database. The database contains 40 video sequences, which contain 10 different subjects with 4 sequences for every subject. Moreover, one gait cycle is determined manually, and we extracted gait features from 3 regions: whole region, upper region, and lower region in average image.

We carried out two experiments as follows: (i) person identification of simulated walking images, (ii) person identification of actual walking images.

4.1 Person identification of simulated walking images,

In the first experiment, we built the database with virtual images captured from 74 different viewpoints, which contain 37 different azimuth angles (every 5 from 0 to 180) and 2 different elevation angles (20 and 25). As test data sets, we used virtual images from 7 different viewpoints (azimuth angle: 0, 30, 60, 90, 120, 150, 180, elevation angles: 20) for each subject as shown in Fig. 7. Here, each test data set was not included in the database (e.g. in case that a test data set is a subject at a viewpoint of 0 degree, the features of the subject is not included in the training data sets.).

Table 1 shows the correct classification rate (CCR) with respect to different viewpoints of the test data sets. From these results, we can see that the proposed method can identify people robustly against the walking direction change.

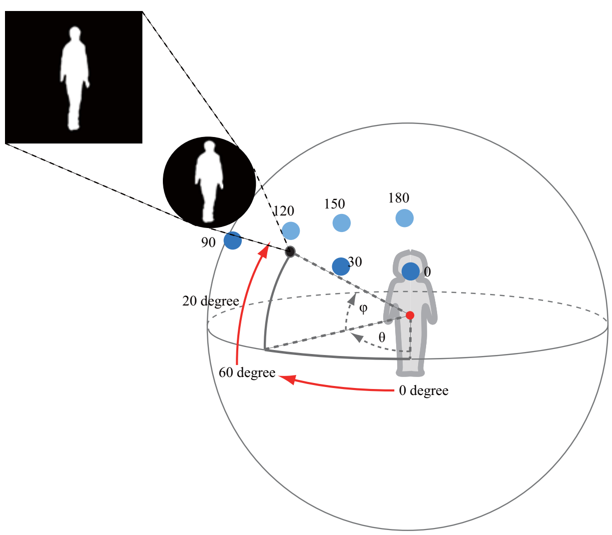


Figure 7. Virtual viewpoints

4.2 Person identification of actual walking images

In the next experiment, we built the database in the same way with the previous experiment, and we used actual images as test data sets, which consists of 40 video sequences containing 5 different subjects with 2 sequences for every subject. Test data sets were captured from 4 different cameras (camera 1 ∼ 4) as shown in Fig. 8, and these cameras were placed around 0, 45, 90, and 135 degree against the walking direction.

Table 1. Correct classification rate with simulated images [%]

Virtual viewpoints [degree]						
0	30	60	90	120	150	180
97.5	90.0	100.0	100.0	100.0	97.5	100.0

Table 2 shows the correct classification rate (CCR) with respect to different cameras, and Table 3 shows the classification result and estimated angle of the virtual viewpoint of each test data set. In Table 3, A ∼ E show the subjects and the number following the subject shows the sequence number. In this experiment, cameras were placed at 0, 45, 90, 135 degree against the walking direction, but these degrees can be changed due to the distance between the camera and the subject. We think this can be the reason estimated angles of camera 3 are varied from 80 to 100. Moreover, according to the results of camera 1 some of estimated angles are around 180, and we think this is because silhouettes captured from both a camera at 0 degree and a camera at 180 degree are almost the same.

Experimental results with actual images are worse than those with simulated images, and one of reasons is that reconstructed 3D model doesn't coincide the real one. Figure 9 show an actual image and a virtual image captured from the same position with the actual camera position.

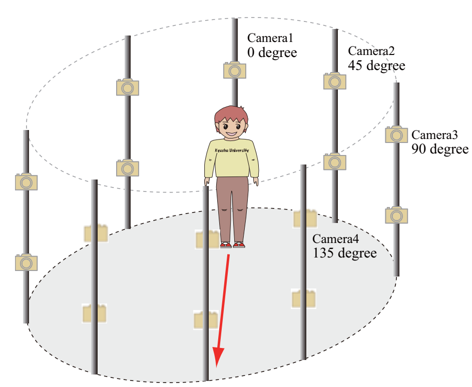


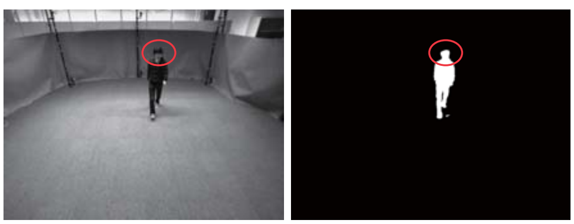
Figure 8. Cameras of test data sets

Table 3. Estimated viewpoints and results of classification (T: True, F: False).

	camera 1(0)		camera 2(45)		camera 3(90)		camera 4(135)	
Targets	angle[degree]	result	angle[degree]	result	angle[degree]	result	angle[degree]	result
A-1	15	F	135	T	80	T	135	T
A-2	15	T	130	T	80	T	135	T
B-1	70	F	145	F	95	T	135	F
B-2	70	F	55	T	100	T	150	T
C-1	5	T	135	F	90	F	80	F
C-2	175	F	140	F	90	T	60	F
D-1	0	F	130	F	90	T	130	F
D-2	180	T	130	F	90	T	130	T
E-1	5	F	80	T	90	T	140	F
E-2	180	F	80	F	90	T	140	F

Table 2. Correct classification rate with actual images [%]

Test cameras			
camera 1 (0 degree)	camera 2 (45 degree)	camera 3 (90 degree)	camera 4 (135 degree)
30	40	90	40



(a) An actual image (b) A synthesized image

Figure 9. Actual image and synthesized image from 3D model

5 Conclusion

We proposed in this paper a spatio-temporal 3D gait database and a view independent person identification method from gait. From experiments with simulated images, we showed that the proposed method identified people with high correct classification rate robustly against the walking direction change. Moreover, from experiments with actual images we showed that a target's walking direction could be estimated and the target was identified with high correct classification rate when one walked in parallel to the camera. Future research will include precise 3D model reconstruction.

6 Acknowledgment

This research was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B) (21700199).

References

[1] J. Acquah, M. Nixon, and J. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24:2175–2183, 2003.

[2] I. Bouchrika and M. Nixon. People detection and recognition using gait for automated visual surveillance. *Proc. IEE Inter. Symp. Imaging for Crime Detection and Prevention*, 2006.

[3] D. Cunado, M. Nixon, and J. Carter. Automatic extraction and description of human gait models for recognition purposes. *CVIU*, 90(1):1–41, 2003.

[4] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167–17, 1993.

[5] J. Flusser, T. Suk, and B. Zitova. *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons Ltd., Erewhon, NC, 2009.

[6] Y. Iwashita and R. Kurazume. Person identification from human walking sequences using affine moment invariants. *Proc. IEEE Int. Conf. Robotics and Automation*, pages 436–441, 2009.

[7] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 143–150, 2003.

[8] S. Lee, Y. Liu, and R. Collins. Shape variation-based frieze pattern for robust gait recognition. *Proc. CVPR*, 2007.

[9] W. Martin and J. Aggarwal. Volumetric description of objects from multiple views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(2):150–158, 1983.

[10] K. Sugiura, Y. Makihara, and Y. Yagi. Omnidirectional gait identification by tilt normalization and azimuth view transformation. *Proc. of the IEEE Workshop on OMNIVIS*, 2008.