

Expanding gait identification methods from straight to curved trajectories

Yumi Iwashita, Ryo Kurazume
Kyushu University
744 Motoooka Nishi-ku Fukuoka, Japan
yumi@ieee.org

Koichi Ogawara
Wakayama University
930 Sakaedani Wakayama, Japan

Abstract

Conventional methods of gait analysis for person identification use features extracted from a sequence of camera images taken during one or more gait cycles. An implicit assumption is made that the walking direction does not change. However, cameras deployed in real-world environments (and often placed at corners) capture images of humans who walk on paths that, for a variety of reasons, such as turning corners or avoiding obstacles, are not straight but curved. This change of the direction of the velocity vector causes a decrease in performance for conventional methods. In this paper we address this aspect, and propose a method that offers improved identification results for people walking on curved trajectories. The large diversity of curved trajectories makes the collection of complete real world data infeasible. The proposed method utilizes a 4D gait database consisting of multiple 3D shape models of walking subjects and adaptive virtual image synthesis. Each frame, for the duration of a gait cycle, is used to estimate a walking direction for the subject, and consequently a virtual image corresponding to this estimated direction is synthesized from the 4D gait database. The identification uses affine moment invariants as gait features. Experiments using the 4D gait database of 21 subjects show that the proposed method has a higher recognition performance than conventional methods.

1. Introduction

Person identification methods have been used for a wide variety of applications, such as surveillance or use of service robots co-located with humans to provide various services in daily life. Gait is one of biometrics that do not require interaction with a subject and can be performed from a distance [11] [15] [12]. Person identification methods which extract features from gait images taken by a camera have been used with good results for human identification [6] [2] [14] [9]. However, since image-based gait recognition is sensitive to appearance changes, the correct classification rate gets low in case that the subject's appearance is different from that in the database. One of possible situations in which this problem occurs is the case that the subject's wak-

ing direction in respect to the camera is different from that in the database (e.g. in Fig. 1(a) walking direction (1) of the subject is different from the direction (2) in the database).

To deal with this problem, several methods have been proposed so far. Makihara *et al.* introduced a view transformation model to synthesize virtual viewpoint images from captured images [14]. In this method, the view transformation model is obtained from training datasets of multiple people which were taken from multiple view points. Iwashita *et al.* introduced a 4D gait database consisting of multiple 3D shape models of walking subjects and a method which identified a subject and estimated his walking direction with synthesized virtual viewpoint images from 3D models in the database [7]. Kusakunniran proposed a method to create a View Transformation Model (VTM) from the different point of view using Support Vector Regression (SVR) [10]. These methods have an implicit assumption that people walk straight and their walking direction does not change during one gait cycle (i.e. he does not walk on curved trajectory). However, in reality people walk on curved trajectories for turning a corner or avoiding an obstacle as shown in Fig. 1 (b).

The large diversity of curved trajectories makes the collection of complete real world data in a database infeasible. This change of the direction of the velocity vector of walking people causes a decrease in performance for conventional methods, which assume the walk direction is straight. In this paper we propose a method to identify people walking on curved trajectories. To the best of our knowledge, this is the first time such a method is introduced to deal with the problem of the decrease of the identification performance due to walking on curved trajectories.

The cause of the performance decrease is that, when a subject walks on a curved trajectory, the observation angle ϕ between the walking direction of the subject and direction of the camera to the subject is gradually changed at all frames in one gait cycle as shown in Fig. 1 (b) (hereafter we call this as "local angle change"). This problem occurs in fact even on a straight walk [3] as shown in Fig. 1 (c). Akae *et al.* showed that the local angle changes in one gait cycle affects the performance of gait identification theoretically, especially in case that the distance between a camera

and a subject is small and side view images are captured [3]. Figures 1 (c) and 2 (b) show examples of images from top and actual images, and it is clear that observation angles between the first frame and the last frame in one gait cycle are different. To deal with this problem, Iwashita *et al.* [8] introduced a method for estimating local angles of a subject walking *straight*, and an experiment with gait images was carried out to illustrate the method.

In this paper, we expand this method [8] to estimate local angles of the subject walking *on a curved trajectory*. The proposed method utilizes a 4D gait database consisting of sequential 3D models of multiple people walking straight and adaptive virtual image synthesis. In the proposed method, one first estimates a walking direction of a subject at each image for the duration of one gait cycle. Next, an observation angle at each frame is estimated from estimated walking direction, and a virtual image is synthesized from each 3D model for the duration of one gait cycle in the 4D gait database so that an observation angle of a synthesized image is the same with that of a frame of the subject. Here, a virtual images is synthesized from a 3D model whose phase is the same with that of a frame of the subject, and virtual images are synthesized from all people in the database. Finally, the subject is identified using affine moment invariants extracted from images as gait features. Our experiments show the effectiveness of the proposed method using two types of gait images: (i) images of people walking on curved trajectories. (ii) images of people walking straight (to evaluate the theory introduced by Akae [3]).

This paper is organized as follows. Section 2 describes the details of the proposed person identification method. Section 3 describes experiments performed using the 4D gait database. Section 4 presents our conclusions.

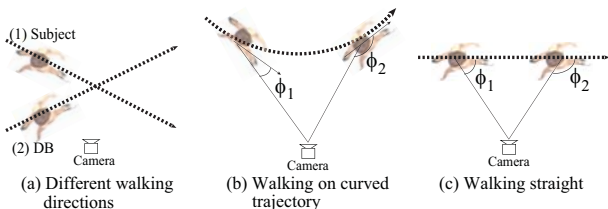


Figure 1. (a) Walking direction change, (b) local angle changes (walking on a curved trajectory), and (c) local angle changes (walking straight).

2. Person identification robust to local angle changes in one gait cycle

In this section, we explain the local angle changes in one gait cycle when a subject walks either on a curved trajectory or straight, and then introduce the proposed method.

2.1. Local angle changes in one gait cycle

The local angle at each frame in one gait cycle is defined with the azimuth angle $\phi_{P_n}^a$ and elevation angle $\phi_{P_n}^e$ at each

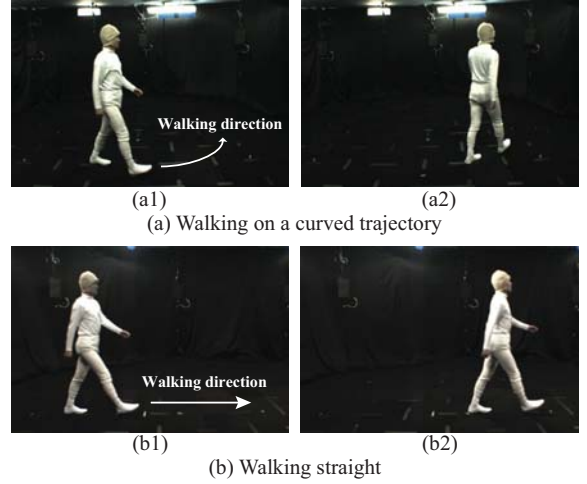


Figure 2. Walking on a curved trajectory and straight. (a1) and (b1) are the first frames in one gait cycle, and (a2) and (b2) are the last frames.

position P_n in one gait cycle as shown in Fig. 3(a). When the subject walks on a curved trajectory as shown in Fig. 3(a), the local angle varies gradually from frame to frame; this deteriorates the performance of gait identification by conventional methods, which have an implicit assumption that the subject walks straight.

When the subject walks straight and the azimuth angle is around ± 90 degree and the elevation angle is small (front/back view images are captured), the local angle change between frames in one gait cycle is small. However, if the azimuth angle is small and camera's position is close to the subject, then the local angle change between frames in one gait cycle cannot be ignored.

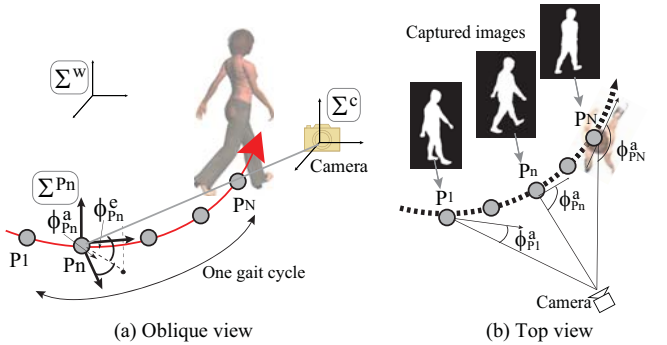


Figure 3. Local angle changes in one gait cycle.

2.2. The proposed person identification method

In this section we describe the details of the proposed method. To summarize, the main steps of processing are as follows.

Step 1 One first estimates a walking direction of a subject at each image for the duration of one gait cycle as

shown in Fig. 3.

Step 2 An observation angle at each frame is estimated from estimated walking direction, and a virtual image is synthesized from each 3D model for the duration of one gait cycle in the 4D database (Fig 4) so that an observation angle of a synthesized image is the same with that of a frame of the subject as shown in Fig. 5. Here, positions (P_1, P_n, P_N) and angles ($\phi_{P_1}^a, \phi_{P_n}^a, \phi_{P_N}^a$) correspond to those in Fig. 3. Virtual images are synthesized from all people in the database

Step 3 Affine moment invariants are extracted as gait features from captures images of a subject and synthesized images, respectively, and the subject is identified.

The 4D gait database includes time-series 3D models of all subjects which were built in advance. The following offers more details.

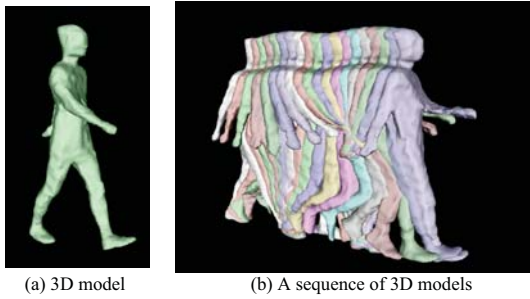


Figure 4. Examples of 3D models in the database.

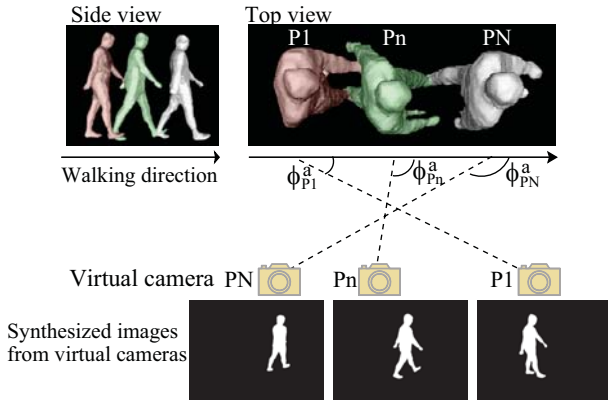


Figure 5. Examples of synthesized virtual images.

2.2.1 Estimation of local angles

In this section we explain a method for estimation of the local angle. Estimating the local angle at each frame in one gait cycle is equivalent to estimating rotation and translation matrices (${}^cR_{P_n}$ and ${}^cT_{P_n}$) from a coordinate system of each foot position Σ^{P_n} to a coordinate system of the camera Σ^c as shown in Fig. 6 (a) (${}^cA_{P_n} = [{}^cR_{P_n} \mid {}^cT_{P_n}]$). To estimate

rotation and translation matrices at each frame in one gait cycle, we estimate foot position and walking direction in 3D space from each captured image. We assume here that the external and internal camera parameters and the position of the floor are known.

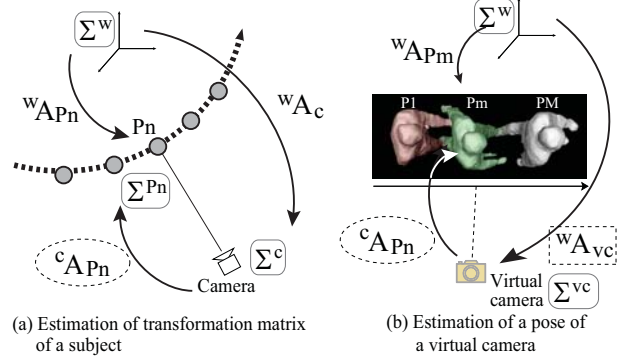


Figure 6. Estimation of transformation matrix of (a) a subject and (b) a virtual camera.

A background subtraction method is applied to captured images to extract silhouette areas. The foot position in the images are estimated as follows: (1) principal component analysis is applied to a silhouette area of each image, and (2) two intersection points of the first principal component and the convex hull of the silhouette area are calculated, and the lower point along the first principal component is estimated as the foot position as shown in Fig. 7.

Next, foot positions ${}^wT_{P_n}$ ($0 \leq n < N$, N is the number of frames in one gait cycle) in 3D space are determined by projecting the estimated foot positions on 2D images onto the floor. Here, ${}^wT_{P_n}$ is a transformation matrix from a coordinate system of each foot position Σ^{P_n} to that of 3D space Σ^w . There are several methods to estimate the walking direction, and one of them is to estimate walking direction at each foot position from difference of the current foot position and its previous one. However, for example when people walk straight, in general they swing from side to side. This swinging can be considered as outliers, and the walking direction calculated by the above method may be different from the real walking direction. Thus, to decrease the influence of outliers, we apply the least square method to foot positions of one gait cycle. In case the subject walks on a curved trajectory, the walking trajectory is estimated by fitting a 2D polynomial to the foot positions of one gait cycle. In case the subject walks straight, the walking trajectory is estimated by fitting a line to the foot positions. Here, to distinguish the walking direction straight and on curved trajectories, at first we fit a line to the foot positions. Then if the sum of distances between the fitted line and foot positions is large, we consider the subject walks on curved trajectory. Finally, the walking direction at each foot position in 3D space is estimated from a differential of the fitted function.

From the estimated walking direction in 3D space and the normal direction of the floor, a rotation matrix ${}^w R_{P_n}$ is determined. Finally the rotation matrix ${}^c R_{P_n}$ from a coordinate system of each foot position Σ^{P_n} to that of the camera Σ^c and the translation matrix ${}^c T_{P_n}$ are calculated as ${}^c R_{P_n} = {}^c R_w {}^w R_{P_n}$ and ${}^c T_{P_n} = {}^c R_w ({}^w T_{P_n} - {}^w T_c)$ as shown ${}^c A_{P_n}$ in Fig. 6 (a). Here, ${}^c R_w$ and ${}^w T_c$ are obtained from external camera parameter.

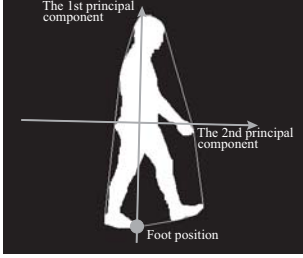


Figure 7. Foot position detection.

2.2.2 Synthesis of virtual images from 3D models

The proposed method relies on a 4D gait database consisting of time-series 3D models of walking subjects which are measured using multiple cameras [8]. The details of the database are explained in Section 3. Figure 4 shows examples of 3D models of a walking person in the database. Note that the database consists of 3D models of multiple people who walked straight. To synthesize virtual viewpoint images, at first the foot position ${}^w T_{P_m}$ ($0 \leq m < M$, M is the number of models in one gait cycle) of each 3D model of the walking person in the database is estimated as the intersection point of the floor and the first principal component, which is calculated by applying the principal component analysis to each 3D model. Then the walking direction of the person is estimated from a line fitted to the foot positions by the least square method, and a rotation matrix ${}^w R_{P_m}$ at each foot position is calculated. Finally, as shown in Fig. 6 (b) the external parameter (${}^w A_{vc}$) of a camera at a virtual viewpoint is calculated from ${}^c R_{P_n}$, ${}^c T_{P_n}$, ${}^w R_{P_m}$, and ${}^w T_{P_m}$, and a virtual viewpoint image is synthesized from a 3D model $Model_m$. Here, we assume that the phase of the first frame of one gait cycle of the subject is the same with that of each person in the database. But the number of frames of the subject in one gait cycle may be different from that of a person in the database, so in general a phase of each frame of the subject should be aligned to that of a synthesized image. In this paper, we align them by $m = \frac{m}{N} \times M$ as a linear solution. Virtual images are synthesized from 3D models at all people in the database with respect to each subject.

2.2.3 Extraction of gait features and person identification

For feature extraction, at first a silhouette area is scaled to a uniform height, set to 128 pixels, and the average image $I^{average}$ from images of one gait cycle is defined by $I^{average}(x, y) = \frac{1}{L} \sum_{l=1}^L I(x, y, l)$ (L is either N for an average image of a subject or M for an average image of synthesized images of each person in the database). Figure 8 shows an example of average images. Then, affine moment invariants are extracted from the average image as gait features [9]. In [9], Iwashita showed that affine moment invariants had the same discrimination capability with GEI [6] and DFT [14] even though the number of features of affine moment invariants is much smaller than those [6] [14].

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

$$\mu_{pq} = \sum \sum_{(x,y) \in O} (x - x_g)^p (y - y_g)^q I(x, y). \quad (1)$$

Here, x_g and y_g are the center of the object. In our method we used thirty affine moment invariants $I = \{I_1, I_2, \dots, I_{30}\}$, and we show two of them [4] [5].

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

In this paper, we divide each average image into K multiple areas as shown in Fig. 8 to obtain features from local areas, and extracted gait features from each area. Here, K is the number of divided areas.

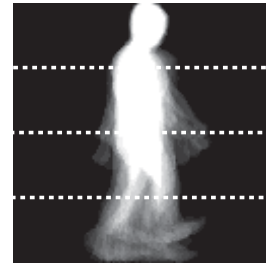


Figure 8. An example of average images (K=4).

2.2.4 Person identification

In the proposed method, we use the nearest neighbor as the classifier. In the training phase, we extract gait features by

the affine moment invariants from synthesized virtual images, and then build a database. Then in the identification phase, gait features are extracted from silhouette images of a subject and the subject is identified by the classifier.

Shakhnarovich *et al.* also utilized 3D models reconstructed with multiple cameras and synthesized images at a virtual camera from 3D models [13]. However, they used multiple cameras for capturing images of subjects and synthesize images from side view only. On the other hand, our proposed method uses only one camera for capturing images of subjects and is able to synthesize images from arbitrary view points.

3. Experiments

This section shows results of person identification experiments using a database which includes the 4D gait database and 2D images of people walking on curved trajectories and straight as test datasets. This database comprises of 21 subjects, with 4 sequences for each subject. All people walked straight as shown in the dashed line (2) in Fig. 9. Multiple 3D models were reconstructed by the visual hull technique with 16 cameras placed in a studio as shown in Fig. 9. We utilized sequential 3D models of all subjects as training datasets and 2D images captured from a camera as test datasets.

Test datasets consist of images of two ways of walking; (*walki*) walking on a curved trajectory and (*walkii*) walking straight. For capturing images of people walking on curved trajectories as shown the dashed line (1) in Fig.9 (*walki*), we selected camera A. The trajectory is part of a circle with radius 1.5 m. The test datasets consist of 21 subjects with 1 sequence. Each gait sequence consists of 15 ~ 20 images. The height of the camera is approximately 1.2 [m], and the distance from each camera to the center of the studio is 3.5 [m].

For test datasets of people walking straight (*walkii*), we utilized captured images taken from a camera (either camera A or camera B). People walked straight as shown the dashed line (2) in Fig.9, and camera A and camera B captured frontal view gait images and side view gait images, respectively. The test datasets for each camera consist of 21 subjects with 4 sequences. Note that for synthesizing virtual images of a subject we eliminated 3D models created from gait images of the same sequence with that of the subject.

There are gait databases publicly available [1], which are for changes of walking direction, and these databases include 2D gait images captured by multiple cameras. However, since these databases does not include information of camera parameters, 3D models of walking people cannot be obtained. Besides, these databases contain images of people walking only *straight*. Thus we do not use these databases in our experiments.

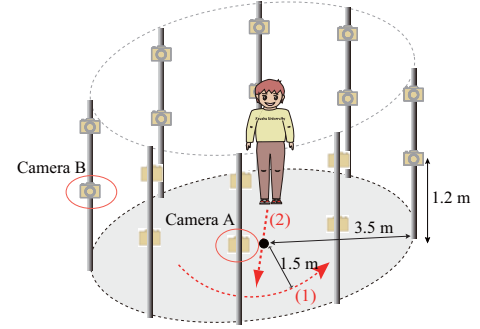


Figure 9. Experimental setting.

3.1. Identification of people walking on curved trajectories

In the first experiments, we applied the proposed method to gait images of people walking on curved trajectories. In the experiments, we applied two methods: (i) the proposed method and (ii) the conventional method [8]. Figure 10(a) shows examples of captured images of a subject, and Figs. 10(b) and (c) show those corresponding synthesized images by the propose method and the conventional method [8], respectively. From these results, it is clear that the proposed method could synthesize similar silhouette of actual images.

To extract gait features, we separate each average image into 1, 2, and 4 areas ($K=1, 2,$ and 4), and we combined gait features from all areas for identification. The correct classification rate by the conventional method [8] and the proposed method were 14.3 % and 71.4 % with sixteen affine moment invariants, which showed high performance in the experiments in the next section. The proposed method showed much higher performance than the conventional method. Moreover, to show the effectiveness of the affine moment invariants, we extracted gait features from synthesized images with GEI [6]. Here, to compare the discrimination capabilities of the affine moment invariants and GEI, we didn't apply any dimensionality reduction methods to GEI. The correct classification rate was 28.6 %, which was much lower than the affine moment invariants. The reason is as follows. In both methods using affine moment invariants and GEI, features are extracted from an average image, and each pixel value of this image can be affected due to slight error of the estimation of the walking direction. Moreover, as we explain in the next paragraph, since the subject walked on a small circle, the number of frames in one gait cycle is less than that in the database. So this can also change pixel values of the average image. In GEI gait feature is extracted at each pixel, so this may be sensitive to the change of pixel values. On the other hand, in the affine moment invariants, gait features, which represents the shape and pixel values of the area, are extracted at each area, so this is robust to the change of pixel values compared with GEI.

When we checked captured images of people walking on curved trajectories, we noticed that some of subjects changed their way of walking due to the small circle. There are some changes which affected their way of walking, such that (i) the length of stride of a subject got short compared with that of the subject walking straight, when the subject walked on a small circle, (ii) the distance between the subject's left and right legs in the frontal plane got bigger or smaller than that of the subject walking straight. Figure 11 shows an example of captured images in the test datasets, an image of the same person from the database with Fig.11(a), and synthesized image of Fig.11(a) from the 3D model of Fig.11(b) by the proposed method. In Figures 11(a) and (b) the person was in the same phase, but it is clear the length of stride in Fig.11(a) is shorter than that in Fig.11(b). In test datasets of 21 people, 3 of them changed their lengths of stride bigger than those of the rest of people. When we removed these 3 people from the test datasets, the correct classification rates by the conventional method and the proposed method were improved to 16.7 % and 83.3 %, respectively.

From above results, the change of their way of walking affected the discrimination capability of gait features from average images. Near term is as follows. When a subject walks, his appearance may be similar to the one in the database even if his way of walking is different from that in the database, so we utilize features from these images for classification.

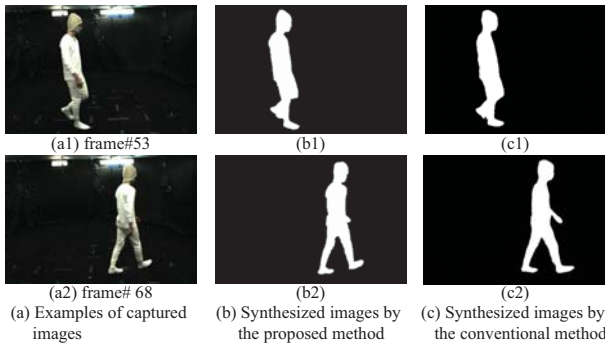


Figure 10. Examples of actual images of a subject walking on a curved trajectory and synthesized virtual images of the subject by the proposed method and the conventional method [8].

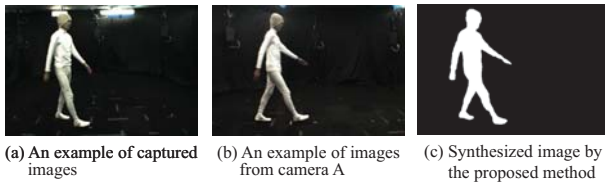


Figure 11. (a) An example of actual images of a subject walking on a curved trajectory, (b) an example of captured images of the same person walking straight with (a), and (c) synthesized image of (a).

3.2. Identification of people walking straight

In the last experiments, we applied the proposed method to images of people walking straight. Figure 12(a1) and (a2), which were captured from camera A show the first frame and the last frame of one gait cycle, respectively, and Fig.12(b) show those corresponding synthesized virtual viewpoint images. Here, for the purpose of visualization of the effectiveness of the proposed method, we utilized 3D models from images of the same sequence with that of the test dataset. From the results of Fig. 12, the proposed method synthesized similar images to the actual images compared with the conventional method.

To evaluate the effectiveness of the proposed method, we applied one of conventional methods [7] to the same test datasets. This conventional method [7] utilizes a fixed local angle (e.g. a local angle at the center position in one gait cycle) to synthesize virtual images of one gait cycle, so a synthesized image at a frame with different local angle from the fixed local angle is different from actual image. This method extracted gait features from average images by affine moment invariants in the same way with the proposed method. Figure 12(c) shows synthesized images which correspond to Fig.12(a).

Figures 13(a) and (b) show examples of captured images from camera B and synthesized images by the proposed method, respectively. These results show that the synthesized images are similar to actual images, and especially in Fig.13(b1) the subject's arm was separated from one's body like the one in Fig.13(a1). The subject's arm in the synthesized image by the conventional method (Fig. 13(c1)) was connected to one's body. From these results, the images synthesized by the proposed method are more similar to captured images.

Figure 14 show the results of correct classification rates from camera A by the proposed method and the conventional method. The highest correct classification rates by the proposed method and the conventional method were 90.5 % and 86.9 %, respectively. Figure 15 shows the results of correct classification rates from camera B by the proposed method and the conventional method with respect to the change of number of affine moment invariants. The highest correct classification rates by the proposed method and the conventional method were 85.7 % and 79.8 %, respectively.

Akai *et al.* showed that the local angle changes in one gait cycle affects the performance of gait identification especially in case that side view images are captured [3]. From above results, when local angle changes are small (camera A), the results obtained using the proposed method are only slightly better than those by the conventional method, due to little local angle changes. However, when local angle changes are big (camera B), the proposed method showed better results than those by the conventional method. These results support the correctness of [3].

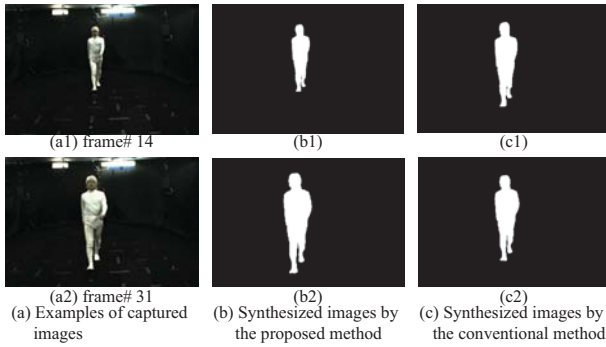


Figure 12. Comparison of virtual images by the proposed method and the conventional method [7] (camera A).

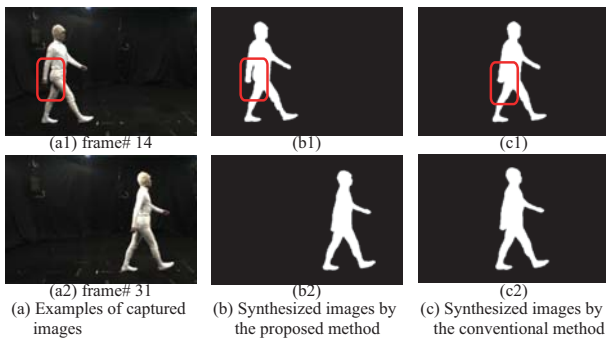


Figure 13. Comparison of virtual images by the proposed method and the conventional method [7] (camera B).

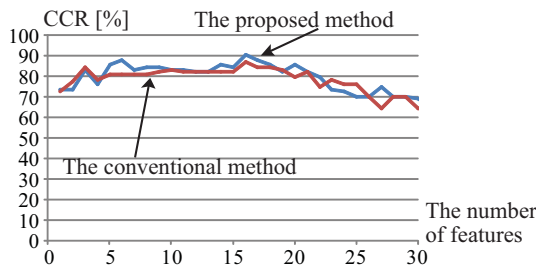


Figure 14. Correct classification rates (camera A).

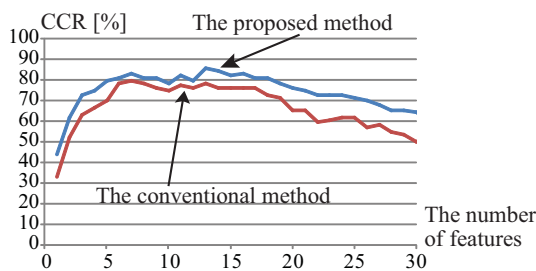


Figure 15. Correct classification rates (camera B).

4. Conclusions

We proposed in this paper the method to identify people walking on curved trajectories by utilizing a 4D gait database and adaptive virtual images synthesis. In this

method, we estimated a walking direction of a subject in 3D space from foot positions of the subject, and we estimated in one gait cycle an observation angle between the walking direction of the subject and direction of the camera to the subject at each frame. Next, a virtual image based on local angle at each position was synthesized from 3D models of all people in the database. Then the subject was identified by gait features extracted by affine moment invariants. We carried out experiments with the 4D database, and showed the effectiveness of the proposed method.

References

- [1] Casia gait database, <http://www.sinobiometrics.com>.
- [2] J. Acquah, M. Nixon, and J. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24:2175–2183, 2003.
- [3] N. Akae, Y. Makihara, and Y. Yagi. The optimal camera arrangement by a performance model for gait recognition. *the 9th IEEE Conf. on Automatic Face and Gesture Recognition*, pages 292–297, 2011.
- [4] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167–174, 1993.
- [5] J. Flusser, T. Suk, and B. Zitova. Moments and moment invariants in pattern recognition. *Wiley & Sons Ltd.*, 2009.
- [6] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. PAMI*, 28(2):316–322, 2006.
- [7] Y. Iwashita, R. Baba, K. Ogawara, and R. Kurazume. Person identification from spatio-temporal 3d gait. *Int. Conf. on Emerging Security Technologies*, pages 30–35, 2010.
- [8] Y. Iwashita, R. Baba, K. Ogawara, and R. Kurazume. Person identification from spatio-temporal volumes. *mW/i*, 2011.
- [9] Y. Iwashita and R. Kurazume. Person identification from human walking sequences using affine moment invariants. *IEEE Int. Conf. Robotics and Automation*, pages 436–441, 2009.
- [10] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Support vector regression for multi-view gait recognition based on local motion feature selection. *Computer Vision and Pattern Recognition (CVPR)*, pages 974–981, 2010.
- [11] M. Nixon, T. Tan, and R. Chellappa. Human identification based on gait. *International Series on Biometrics Springer*, 2011.
- [12] A. Roy, S. Sural, and J. Mukherjee. Gait recognition using pose kinematics and pose energy image. *Signal Processing*, 92(3), 2012.
- [13] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. *Computer Vision and Pattern Recognition*, 1:439–446, 2001.
- [14] K. Sugiura, Y. Makihara, and Y. Yagi. Gait identification based on multi-view observations using omnidirectional camera. *Asian Conf. on Computer Vision*, 1:452–461, 2007.
- [15] E. Zhang, Y. Zhao, and W. Xiong. Active energy image plus 2dlpp for gait recognition. *Signal Processing*, 90(7):2295–2302, 2010.