

TECHNICAL NOTE

Open Access



The OU-ISIR Large Population Gait Database with real-life carried object and its performance evaluation

Md. Zasim Uddin^{1*}, Thanh Trung Ngo¹, Yasushi Makihara¹, Noriko Takemura², Xiang Li^{1,3}, Daigo Muramatsu¹ and Yasushi Yagi¹

Abstract

In this paper, we describe the world's largest gait database with real-life carried objects (COs), which has been made publicly available for research purposes, and its application to the performance evaluation of vision-based gait recognition. Whereas existing databases for gait recognition include at most 4007 subjects, we constructed an extremely large-scale gait database that includes 62,528 subjects, with an equal distribution of males and females, and ages ranging from 2 to 95 years old. Moreover, whereas existing gait databases consider a few predefined CO positions on a subject's body, we constructed a database that contained unconstrained variations of COs being carried in unconstrained positions. Additionally, gait samples were manually classified into seven carrying status (CS) labels. The extremely large-scale gait database enabled us to evaluate recognition performance under cooperative and uncooperative settings, the impact of the training data size, the recognition difficulty level of the CS labels, and the possibility of the classification of CS labels. Particularly, the latter two performance evaluations have not been investigated in previous gait recognition studies.

Keywords: Gait database, Extremely large scale, Carried object, Carried object detection and classification, Performance evaluation

1 Introduction

Gait refers to the walking style of an individual, and can be used as a behavioral biometric [28]. Compared with traditional biometric features, such as DNA, a fingerprint, face, and iris, gait has many unique advantages. The key advantage is that gait can be used to recognize an individual at a distance from a camera without his/her cooperation, even for a relatively low-resolution image sequence [36] and low frame rate [25]. Therefore, gait has the potential to be applied in many applications, such as access control, surveillance, forensics, and criminal investigations from footage from CCTV cameras installed in a public or private space [4, 16, 19]. Recently, gait has been used as a forensic feature, and there has already been a conviction that has resulted from gait analysis [37].

However, gait recognition has to overcome some practical issues because of circumstances defined as covariates, such as view, clothing, shoes, carried object (CO), environmental context, aging, or mental condition [30, 40]. These covariates should be fully studied for further progress and the development of a practical and robust gait recognition algorithm. To overcome these issues, a common gait database that considers the above covariates is essential. Among the aforementioned covariates, CO is one of the most important because people often need to carry objects in their daily lives, such as a handbag, briefcase on the way to work, or multiple bags after shopping.

There are some existing gait databases in the research community that consider COs. However, they contain a limited number of subjects and few predefined COs, and they lack information about the positions and types of COs. For example, CASIA gait dataset B [40] is composed of 124 subjects and considers a bag as a CO, where the bag was selected by a subject from a predefined set

*Correspondence: zasim@am.sanken.osaka-u.ac.jp

¹The Institute of Scientific and Industrial Research, Osaka University, Osaka 567-0047, Japan

Full list of author information is available at the end of the article

containing a knapsack, satchel, and handbag. Similarly, the USF dataset [30] is composed of 122 subjects and considers a briefcase as a CO; thus, there are at most two options for a CO available, that is, with or without a briefcase. Recently, a large dataset, the OU-ISIR Gait Database, Large Population Dataset with Bag, β version, which contains 2,070 subjects with various COs, was introduced in [22]; however, it does not include detailed information about COs.

With the growing data science trend, we always need a large-scale dataset to efficiently solve a problem. Recently, many sophisticated machine learning techniques, such as deep learning (DL), have been developed, and they require a large number of training samples because more data are more important than a better algorithm [9]. However, few large-scale databases are available for gait recognition, for example, the OU-ISIR Gait Database, Large Population Dataset [15] and Large Population Dataset with Bag, β version [22], which consider 4,007 and 2,070 subjects, respectively. Although these datasets for gait recognition seem to be sufficient for a conventional machine learning algorithm (e.g., without DL), they are not sufficiently large to efficiently conduct a study using a DL-based approach.

In this study, we first propose an extremely large population gait database with a large variation of CO covariates that will encourage the gait recognition community to deeply research this practical covariate. Second, we provide performance evaluations for gait recognition by employing existing state-of-the-art appearance-based gait representation. The contributions of this paper are summarized as follows:

1. The proposed database¹ is the largest gait database in the world and is constructed from 62,528 subjects with an equal distribution of males and females, and

a wide range of ages. It is more than 15 times the size of the existing largest dataset for gait recognition.

2. In the proposed database, there is no constraint on type, quantity, and position of the CO. We considered any real-life COs that are used in daily life (e.g., handbag, vanity bag, book, notepad, and umbrella) or when traveling (e.g., backpack, luggage, and travel bag). Additionally, the typical position labels of the COs are manually annotated. It would be beneficial to analyze the classification and gait recognition difficulty with respect to these typical position labels.
3. We provide a set of evaluation experiments with benchmark results using state-of-the-art gait recognition algorithms. Particularly, experiments related to COs have not been investigated in previous gait recognition studies.

2 Related work

2.1 Existing gait recognition databases

In this section, we briefly describe the existing major databases for gait recognition, which are summarized in Table 1.

The USF dataset [30] is one of the most widely used gait datasets and captured outdoors under different walking conditions. It is composed of 122 subjects and considers a briefcase as a CO, and as a result, at most two options for samples (i.e., with or without a CO) are available.

The Soton small dataset [27] considers only three types of bags (i.e., handbag, barrel bag, and rucksack) as COs and the subject carries these bags in four ways. Because this dataset contains a larger variation of CO covariates than that of the USF dataset, it can be used for exploratory CO covariate analysis for gait recognition [3].

The TUM-IITKGP [12] dataset contains unique covariates, such as dynamic and static occlusion. Later, TUM-GAID [13] dataset is constructed and it is the first

Table 1 Existing major gait recognition databases

Database	#Subjects	Types of CO	#Possible options for CO positions	Gender balance (male:female)
Soton dataset, small [27]	12	Handbag, barrel bag, rucksack	Four	N/A
USF dataset [30]	122	Briefcase	One	4:1
CASIA dataset, B [40]	124	Knapsack, satchel, handbag	Three	3:1
CASIA dataset, C [34]	153	Bag	One	6:1
CMU Mobo dataset [10]	25	Ball	One	12:1
TUM-IITKGP [12]	35	Backpack	One	N/A
TUM-GAID [13]	305	Backpack	One	3:2
OU-ISIR, LP [15]	4,007	N/A	N/A	1:1
OU-ISIR, LP with Bag, β version [22]	2,070	Unconstrained	Unconstrained	N/A
Proposed	62,528	Unconstrained	Unconstrained	1:1

multi-signal gait dataset to contain audio signals, RGB images, and depth images by Microsoft Kinect.

CASIA dataset B [40] is constructed from 124 subjects with and without a CO, and before capturing the sequences with a CO, each subject chose a bag from a set of the knapsack, satchel, or handbag that he/she liked. As a result, there are at most four options of samples available regarding COs (no bag, knapsack, satchel, and handbag). CASIA dataset C [34] considers only a backpack as a CO, and data was captured from 153 subjects using a thermal infrared camera designed for the study of night gait recognition.

OU-ISIR, LP with Bag, β version [22], is composed of 2,070 subjects and considers unconstrained types and positions of COs. However, information about the status of COs, such as position, change of position within a gait period, and quantity of COs, is unavailable.

To summarize, the aforementioned datasets are unsuitable not only for studying CO covariates but also for taking advantage of modern machine learning (e.g., DL) approaches. By comparing existing databases, the proposed database contains unconstrained variations of COs and the largest number of subjects, which is approximately 200 times larger than the largest existing gait database with COs, that is, TUM-GAID, and 15 times larger than that without COs for gait recognition, that is, OU-ISIR, LP.

We note that there exists a larger gait database [39] that consists of 63,846 subjects. However, this database is only used for age estimation and is not usable for gait recognition because only a single gait energy image (GEI) feature is available for each subject.

2.2 Gait recognition approaches

In gait recognition, the appearance-based approach is dominant, and GEI [11] is the most prevalent and frequently used feature. Furthermore, some modified GEIs have been introduced for robust gait recognition against CO and clothing variation covariates, such as Gait Entropy Image (GEnI) [1], which is computed by calculating the Shannon entropy for every pixel of the GEI; Masked GEI (MGEI) [2], for which gait energies are masked out when gait entropy is smaller than a certain threshold; Gabor GEI [35]; and transformed GEI [18] with a Gabor filter.

Appearance-based features, however, often suffer from large intra-subject appearance changes because of covariates. To gain more robustness, the most popular approach is to incorporate spatial metric learning-based approaches, such as linear discriminant analysis (LDA) [29] and a ranking support vector machine (RankSVM) [7]. Additionally, as a DL-based approach, a convolutional neural network (CNN) [8, 32, 38] is also used for robust gait recognition. Therefore, in this study, we consider

metric learning-based approaches with a GEI feature to evaluate the performance of the proposed database.

3 OU-ISIR large population gait database with carried objects

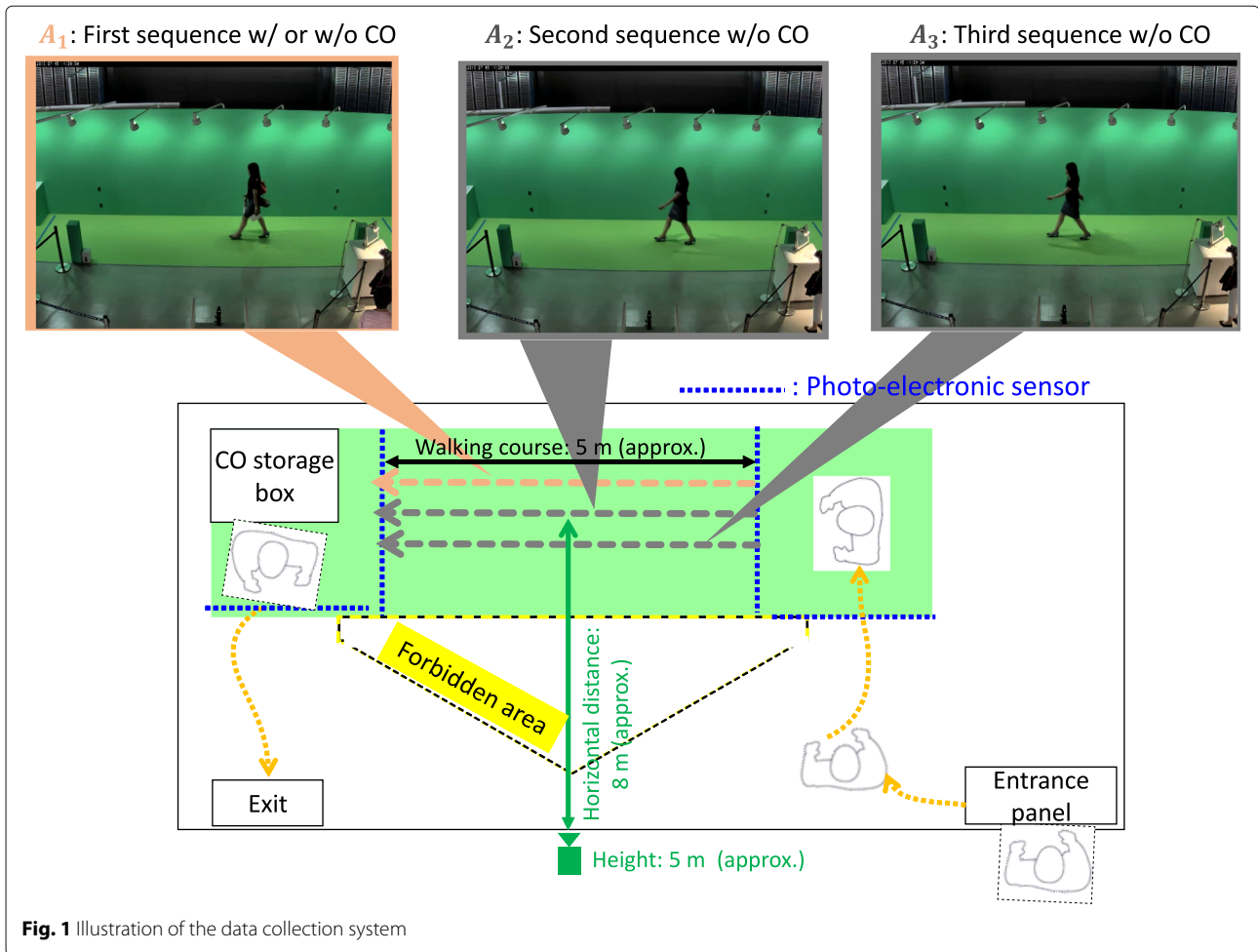
3.1 Capture system

The proposed database was constructed from gait images automatically collected by a gait collecting system called Gait Collector [21]. The gait data were collected in conjunction with an experience-based demonstration of video-based gait analysis at a science museum (Miraikan), and informed consent for purpose of research use was obtained electronically. An overview of the capture system is illustrated in Fig. 1. The camera was set at a distance of approximately 8 m from the straight walking course and a height of approximately 5 m. The image resolution and frame rate were 1280×980 pixels and 25 fps, respectively. The green background panels and carpet were arranged along the walking course for clear silhouette extraction. The camera continuously captured video during the museum opening hours, photo-electronic sensors were used for detecting a subject walking past, and a sequence of a target subject was extracted from the entire video stream.

Each subject was asked to walk straight three times at his/her preferred speed. First, the subject walked to the other side of the course with his/her COs and then placed these items into a CO storage box. Subsequently, he/she walked twice more without COs in the same direction and then picked up the COs and left the walking course. As a result, we obtained three sequences for each subject. The first sequence with or without COs (if he/she did not have COs) is called the A_1 sequence, and the second and third sequences without COs are called A_2 and A_3 sequences, respectively.

3.2 Gait feature generation

To obtain a GEI feature, we performed the following four steps [15]: (1) A silhouette image sequence of a subject was extracted using a chroma-key technique [31] (i.e., removal of the green background area using HSV color space). (2) Then, registration and size normalization of the silhouette images were performed. First, the subject's silhouette images were localized by detecting the top, bottom, and horizontal center (i.e., median) positions. Then, a moving-average filter was applied to these positions. Finally, the sizes of the subject's silhouette images were normalized according to the average positions so that his/her height was 128 pixels. Furthermore, the aspect ratio of each region was maintained, and as a result, we generated the subject's silhouette images of 88×128 pixels. (3) A gait period was determined using normalized autocorrelation [15] of the subject's silhouette image sequence along the temporal axis. (4) A GEI was constructed by averaging the



subject's silhouette image sequence over a gait period. If several gait periods were detected from one walking image sequence, then we chose a GEI that was nearest to the center of the walking course.

3.3 Annotation of the carrying status

Because we did not constrain the subject in terms of the type of CO, or where and how it was carried, it could be carried in a variety of positions and orientations. Thus, it was difficult to categorize the position exactly. For simplicity, we first divided the area in which the COs could be carried into four regions with respect to the human body: side bottom, side middle, front, and back, as shown in Fig. 2. However, some subjects did not carry a CO, some carried multiple COs in multiple regions, and others changed a CO's position within a GEI gait period.

For each GEI, every fourth frame within a gait period was manually checked to annotate the carrying status (CS). As a result, a total of seven distinct labels for the CS were annotated in our proposed database. A summary of the denotation of the CS labels is shown in Table 2 and some examples of CS labels in Fig. 3. Note that, because

only the samples for the A_1 sequence may have contained COs, the annotation process was only applied to the A_1 sequence for each subject.

3.4 Database statistics

Because of the good design of the system, the world's largest database for gait recognition with COs, composed of 62,528 subjects with ages ranging from 2 to 95 years, was constructed. Detailed distributions of the subjects' genders by age group are shown in Fig. 4. The gender distribution was well-balanced for males and females

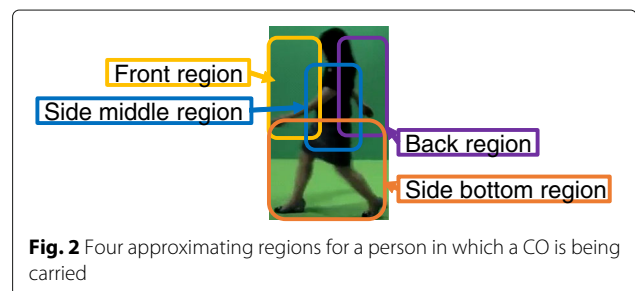


Table 2 Carrying status label

CS label	Explanation
NoCO	No carried object
SbCO	CO(s) being carried in the side bottom region
SmCO	CO(s) being carried in the side middle region
FrCO	CO(s) being carried in the front region
BaCO	CO(s) being carried in the back region
MuCO	COs being carried in multiple regions
CpCO	CO(s) with position being changed from one region to another within a gait period

for each age group, which is a desirable property for the comparison of gait recognition performance between genders [20].

Improper GEIs were excluded manually from the final database if a subject stopped walking for a while at the center of the walking course, changed walking direction before the end of the walking course, continued to carry COs in the A_2 and A_3 sequences, or exited from the capture system after finishing the first sequence, A_1 . As a result, each subject had at most three sequences. We, therefore, obtained a database for publication that included 60,450 subjects for the A_1 sequence, and 58,859 and 58,709 subjects for A_2 and A_3 sequences, respectively.

The distributions of the CS labels are shown in Fig. 5. Most of subjects carried multiple COs in multiple regions (i.e., with MuCO) and the subjects equally liked to carry COs at the front (i.e., with FrCO) and back regions (i.e., with BaCO). Additionally, the subjects equally did not like to carry COs (i.e., with NoCO). Moreover, few subjects changed their CO positions from one region to another (i.e., with CpCO); similarly, few subjects carried COs in the side middle region. Meanwhile, the number of subjects who carried COs in the side bottom region (i.e., with SbCO) was approximately twice as many as those who carried COs in the side middle region (i.e., with SmCO).

4 Performance evaluation

4.1 Overview

These experiments were designed to address a variety of challenges for COs and provided benchmark results for a competitive performance comparison of various algorithms. Specifically, we considered two sets of popular experiments for gait recognition: *cooperative and uncooperative settings* and *impact of the number of training subjects*. Additionally, we designed two more sets of original experimental settings to study the impact of COs: *difficulty level of the CS labels* and *classification of the CS labels*. To the best of our knowledge, they have not been investigated before.

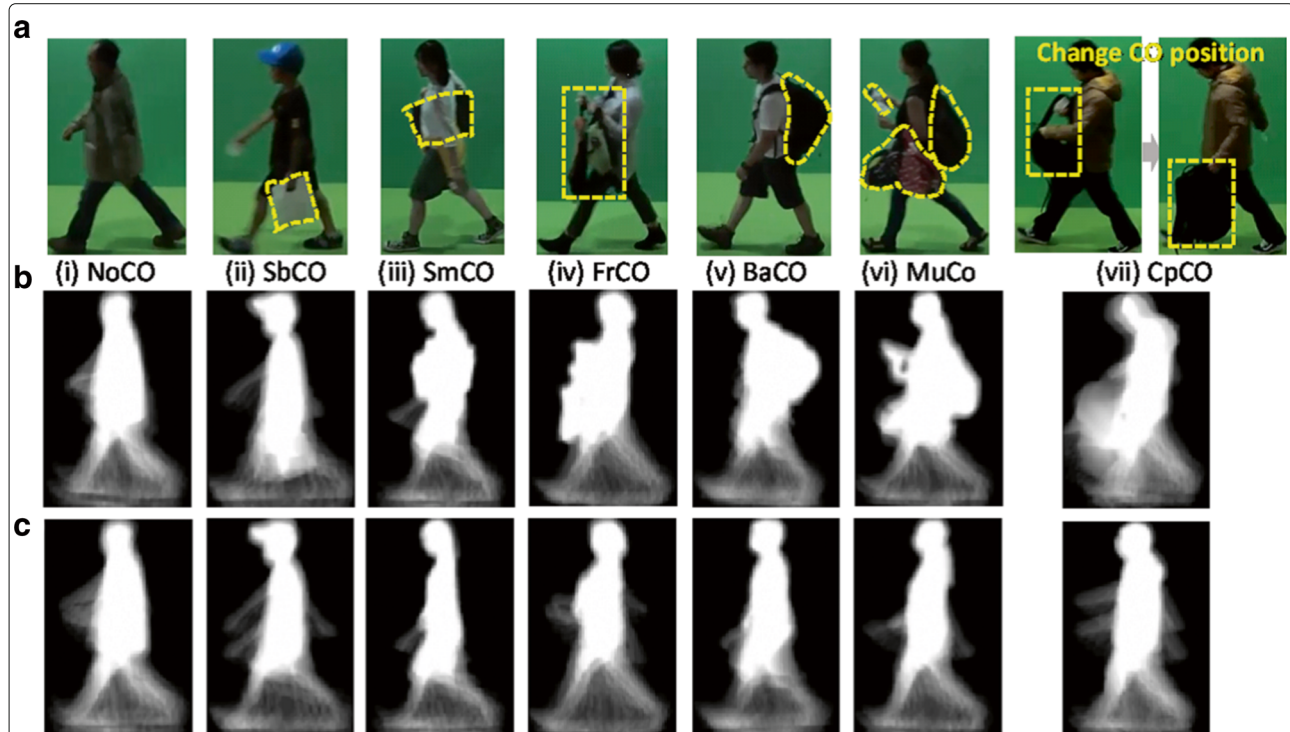
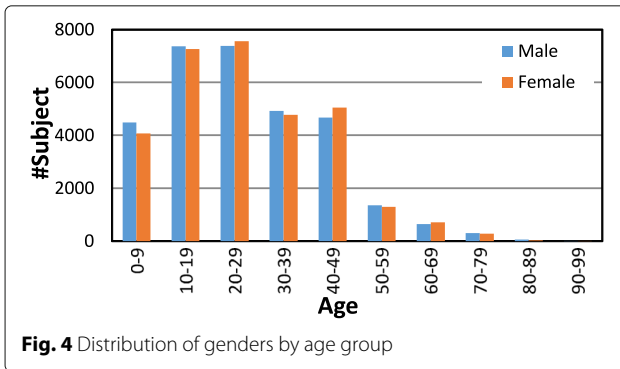


Fig. 3 Examples of CS labels: **a** sample RGB image within a gait period with COs (circled in yellow) in their A_1 sequence; **b** corresponding GEL feature; and **c** GEL feature of the same subject without a CO in another captured sequence (A_2 or A_3), for reference



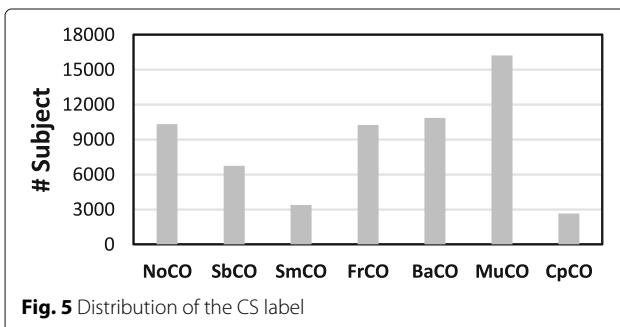
4.2 Evaluation criteria

We evaluated the accuracy of gait recognition in two modes: identification and verification. We used the cumulative matching curve (CMC) for identification and the receiver operating characteristic curve with z-normalization (z-ROC), which indicates the trade-off between the false rejection rate (FRR) of genuine samples and false acceptance rate (FAR) of imposter samples with varying thresholds for verification. Additionally, more specific measures for each evaluation mode were used to evaluate performance: Rank-1 and Rank-5 for identification, and the equal error rate with z-normalization (z-EER), FRR at 1% FAR with z-normalization (z-FRR_{1%}), and area under curve with z-normalization (z-AUC) for verification.

Additionally, we used the correct classification rate (CCR) to evaluate accuracy for the classification of the CS label experiment.

4.3 Benchmarks

There are various state-of-the-art appearance-based methods available for gait recognition in the literature, as mentioned in the Subsection 2.2. We selected seven benchmark methods from the wide variety of appearance-based gait recognition methods to validate the proposed database, which are summarized as follows:



- The first benchmark used the direct matching method [15], which is a non-training-based approach that calculates the dissimilarity using the L_2 distance between two GEIs. The method is denoted by DM in this paper.
- The second benchmark used LDA [29], which is widely exploited in gait recognition [14, 18]. Specifically, we first applied principal component analysis (PCA) to an unfolded GEI feature vector to reduce its dimensions, and subsequently applied LDA to obtain a metric to recognize an unknown sample. The benchmark is denoted by PCA_LDA in the experiment discussions.
- The third benchmark used the gait energy response function (GERF) [18], which transforms GEI into a better discriminative feature. Then, a Gabor filter was applied to the transformed GEI, and LDA was subsequently applied, followed by PCA. The benchmark is denoted by GERF in the experiment discussions.
- A support vector machine (SVM) [6] is a widely used method for multi-class classification. Therefore, we used SVM in a benchmark, with a third-degree polynomial kernel for the classification of the CS labels. The benchmark is denoted by mSVM in the experiment discussions.
- RankSVM [7] is a well-known extension of a SVM that is used for gait recognition in the literature [23, 24, 26]. Therefore, we used RankSVM in a metric learning-based benchmark. In the training phase, we set the positive and negative feature vectors as the absolute difference between the genuine and imposter pair of GEIs, respectively. By considering the computational cost and memory, we selected randomly nine imposter pairs against a genuine pair. The benchmark is denoted by RSVM in the experiment discussions.
- GEINet [32] is based on a simple CNN network architecture for gait recognition, in which one input GEI feature is fed into the network, and the soft-max value from the output of the final layer (fc4), in which the number of nodes is equal to the number of training subjects, is regarded as the probability that the input matches a corresponding subject. The benchmark is denoted by GEINet in the experiment discussions.
- Siamese [8] is also based on CNN network architecture, in which two input GEI features are used to train the two parallel CNN networks with shared parameters for gait recognition [33, 41]. The output of the final layer (fc4) is regarded as a feature vector for each input. A contrastive loss was used for the genuine pair, whereas a so-called hinge loss was used for the imposter pair. Note that, for training the

network, similar to RSVM, we set nine imposter pairs against a genuine pair. The benchmark is denoted by SIAME in the experiment discussions.

4.4 Cooperative and uncooperative settings

In this section, the impacts of the cooperative and uncooperative settings for recognition accuracy are investigated. The implicit assumption for the cooperative setting is that the covariate condition is consistent for all samples in a gallery set. However, it is difficult to collect such data in a real scenario because of the uncooperative and non-intrusive traits of gait biometrics. Therefore, in addition to the cooperative setting, a more natural uncooperative setting was used in which the covariate condition was inconsistent in the gallery set [24].

For the settings, we prepared a subject list that included 58,199 subjects who had a sample in the A_1 sequence and a sample in either the A_2 or A_3 sequences for each subject. Then, the subject list was divided randomly by subject id into two sets: a training set (29,097 subjects) and test set (29,102 subjects) equally for each CS label. Then, the test set was divided into two subsets: gallery set and probe set. For the cooperative setting, we used samples from the A_2 or A_3 sequences (i.e., without COs) in the gallery and the sample from the A_1 sequence was used as a probe. While in the uncooperative setting, samples of each subject were randomly separated into a gallery set and probe set so that the gallery contained a mix of samples that consisted of A_1 and A_2 or A_3 sequences. The training sets for the cooperative and uncooperative settings were prepared in the same manner to reflect the corresponding test sets.

The results for CMC and z-ROC are shown in Fig. 6, and Rank-1, Rank-5, z-FRR_{1%}, z-EER, and z-AUC are shown in Table 3. From these results, the recognition accuracy for the cooperative setting is better than that of the uncooperative setting for most of the benchmarks.

Among the benchmark methods, the non-training-based approach DM achieved the worst performance. Because DM did not apply a technique against the covariate, it was directly affected by the spatial displacement of the corresponding body parts in GEIs caused by the CS difference. By contrast, the accuracy of the training-based approaches was better than that of DM because the dissimilarity metrics were optimized with the training dataset.

Regarding the LDA-based metric learning benchmarks, both PCA_LDA and GEF worked reasonably well and their performances were very similar. However, GEF was slightly better for the uncooperative setting, whereas PCA_LDA was slightly better for the cooperative setting, as shown in Fig. 6 and Table 3. We believe that LDA performed better recognition for both benchmarks by reducing intra-subject appearance variation while increasing inter-subject variations. Furthermore, in GEF, before applying LDA and PCA, a pre-processing technique was performed on GEI, for example, transforming a pixel value for a better discriminative feature. This transformation in GEF was not effective for the cooperative setting; however, it worked well for the uncooperative setting. As a result, the performance of GEF was better for the uncooperative setting.

Regarding RSVM, it is reported in the literature that RankSVM works better in an identification scenario [24, 36] because it focuses more on the relative distance between two classes and considers the probe-dependent rank statistics. However, it did not work well in our setting. We believe the cause of this weak performance was that, as mentioned in Section 4.3, we could only set the number of impostor pairs at nine against a genuine pair, and hence, RankSVM could not effectively maximize the inter-subject variation. This is one of the important disadvantages of the RankSVM method for an extremely large training dataset.

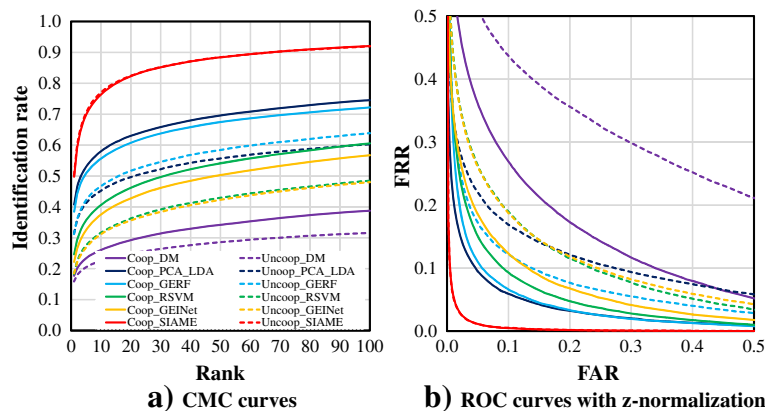


Fig. 6 CMC and ROC curves for cooperative and uncooperative settings. Legend marks are common in all graphs. **a** CMC curves. **b** ROC curves with z-normalization

Table 3 Rank-1/5 [%], z-FRR_{1%}, z-EER [%], and z-AUC [%] for cooperative (Coop) and uncooperative (Uncoop) settings

Benchmark	Rank-1		Rank-5		z-FRR _{1%}		z-EER		z-AUC	
	Coop	Uncoop	Coop	Uncoop	Coop	Uncoop	Coop	Uncoop	Coop	Uncoop
DM	17.7	15.9	23.4	20.5	56.3	68.0	18.5	29.9	10.1	23.2
PCA_LDA	40.8	31.4	53.0	41.3	21.2	34.3	7.4	14.4	2.4	8.0
GERF	38.5	31.2	50.9	42.2	30.6	34.6	8.0	11.4	2.7	5.1
RSVM	24.7	18.3	35.6	27.6	34.1	43.9	9.6	14.7	3.5	8.2
GEINet	22.3	18.5	32.5	26.9	34.8	43.3	11.3	14.7	4.5	7.1
SIAME	49.8	50.3	69.7	70.5	5.7	5.4	2.5	2.4	0.3	0.3

Bold and italic bold fonts indicate the best and second-best benchmarks, respectively

Regarding CNN-based benchmarks, although GEINet did not work well, SIAME achieved the best results with a large margin compared with other benchmarks. We believe the cause of the weak performance for GEINet was that the parameter of the one-input CNN architecture was trained so as to maximize the soft-max of the output layer (fc4) node for the same subject's input GEIs. Therefore, it emphasized minimizing only intra-subject appearance variation. However, only two sample GEIs for each subject were used in these experiments, which was not sufficient to train a good parameter. By contrast, the two-input CNN architecture Siamese in SIAME was trained so that it minimized the variation between the intra-subject and maximized the variation between inter-subject GEIs. Furthermore, there was no accuracy deviation between the cooperative and uncooperative settings for SIAME. We believe that the deep neural network structure of Siamese was sufficiently powerful to manage CO covariates given a very large training dataset.

4.5 Difficulty level of the CS labels

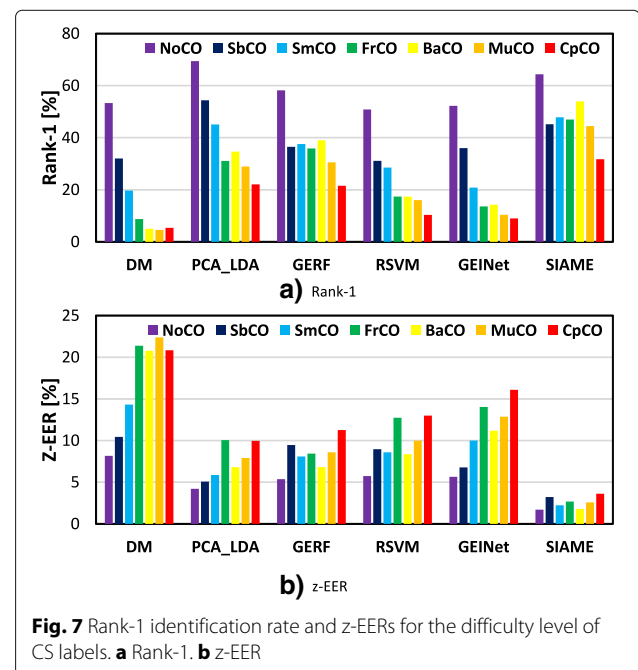
The purpose of this experiment was to analyze the difficulty level of the CS labels based on recognition performance. To analyze the difficulty level, we used the same protocol as the cooperative setting, except the probe set was divided into seven subsets according to the CS label, whereas the gallery set was unchanged for a fair comparison.

The results for the Rank-1 identification rate and z-EERs are shown in Fig. 7. NoCO and CpCO achieved the best and worse labels respectively, whereas the remaining labels (i.e., SbCO, SmCO, FrCO, BaCO, and MuCO) were approximately at the middle difficulty level. We discuss the evaluation results by considering the static shape and dynamic motion of the gait feature.

NoCO was the best label for any benchmark, and this is reasonable because there was no CO between the gallery and probe of the same subject and, as a result, shape and motion were stable.

Regarding the middle-level difficulty labels, the motion and shapes deviated by different amounts. For example,

for SbCO and SmCO, subjects frequently carried small and lightweight COs, which were occluded by the subject's body very often, as shown in Fig. 3. Therefore, the COs did not have much of an impact on the shape. For the case of BaCO, subjects typically carried a large CO, such as a backpack that was secured by two straps that fitted over the shoulders, and thus the position of the CO was fixed and stable within a gait period. However, the large CO heavily affected the shape and posture, as shown in Fig. 3. Similarly for MuCO, subjects typically carried a large backpack-type CO together with other types of COs that were carried in other regions. Although the CO position of the back region was constant, other CO positions were random; thus, GEI samples for MuCO were largely affected not only by shape but also by motion. As a result, the recognition performance of this label was worse than that of BaCO. Regarding FrCO, the subjects typically carried a lightweight object by hand in the front region.



Unlike BaCO, the CO position was not stable, and typically both hands were required to hold the CO in the front region. Therefore, the GEI samples of FrCO were affected slightly by shape and fairly affected by motion.

Regarding CpCO, the CO position was random in any region within a gait period because of the randomly changing position from one region to another. Therefore, GEI samples for CpCO were severely affected by the motion feature, in addition to shape. As a result, CpCO was the most difficult label.

4.6 Impact of the number of training subjects

It is well-known that the performance of a machine learning-based method, particularly modern machine learning, depends on a variety of training samples. In a specific scenario, such as our case, this variety can be expressed by the number of subjects. In this section, the impact of the number of training subjects on recognition performance is investigated.

In the experiment, we chose the cooperative setting of Section 4.4 and selected the best benchmark, that is, the CNN-based benchmark SIAME. Then, we prepared the training set for 100, 200, 500, 1000, 2000, 5000, and 10,000 subjects randomly from the entire training set (29,097), and the test set was unchanged.

The results for Rank-1 identification rate and z-EERs are shown in Fig. 8. The accuracy was better for a larger number of training subjects. For example, z-EER reduced by approximately 13% when the number of training subjects increased from 100 to 29,097, whereas the rank-1 identification rate increased by approximately 44%.

The above results clearly demonstrate the importance of the number of training subjects, and a large database is essential.

4.7 Classification of the CS labels

In previous sections, we presented our evaluations of subject recognition based on gait, and in this section, we evaluate a different recognition problem, that is, the classification of the CS labels based on the gait feature. There

could be numerous applications such as, the detection of suspicious events, such as bag-prohibited area incursion and locating the person with a backpack. However, there is no standard gait-based CO database with available labeling information about the position and type of CO. Thus, most existing work in the gait recognition literature detects a CO using the gait feature [5, 17, 35]; however, they only classify with or without a CO. We believe that to overcome such a limitation, our proposed database can be used as a benchmark database for the detection and classification of CO positions because of the available labeling information of the CO.

To evaluate the performance of the classification of the CS labels, we divided the number of subjects for each label into a training set and test set equally. Because the number of training subjects for each label was not the same, we equalized the number of training subjects for all labels by considering the smallest number of training subjects for a label, that is, for CpCO (1,300 subjects). For each subject, only the GEI of the A_1 sequence was used. Then, we trained the training-based benchmarks using the equalized training set. For testing, each sample of a CS label was matched against all the available samples of the training set. To predict the CS label for each test sample, majority voting was used for mSVM [6] and the mean distance to a class was used for all other benchmarks.

The CCR results of all CS labels for each benchmark are shown in Fig. 9. The confusion matrices for the best and second-best benchmarks, which were SIAME and mSVM, respectively, are shown in Table 4 for all labels as an average accuracy. The classification accuracy for each label was quite different and depended on the benchmark.

Regarding the performance of benchmark methods, SIAME and mSVM consistently worked for each label, as shown in Fig. 9. For SIAME, as already mentioned in Section 4.4, the Siamese network was trained by minimizing the distance between intra-labels and maximizing the distance between inter-labels. Even mSVM used a shallow machine learning approach (i.e., SVM), but it worked well. We believe the cause is that multi-class SVM [6] constructed multiple binary classifiers (e.g., $K(K-1)/2$

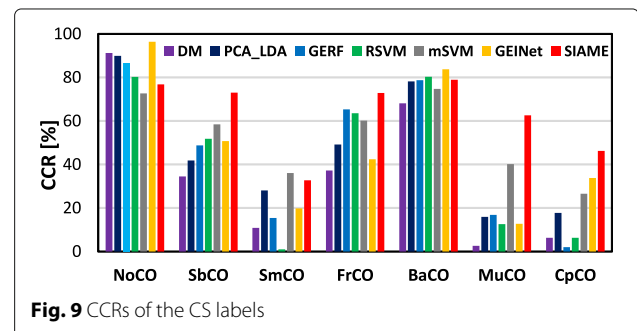
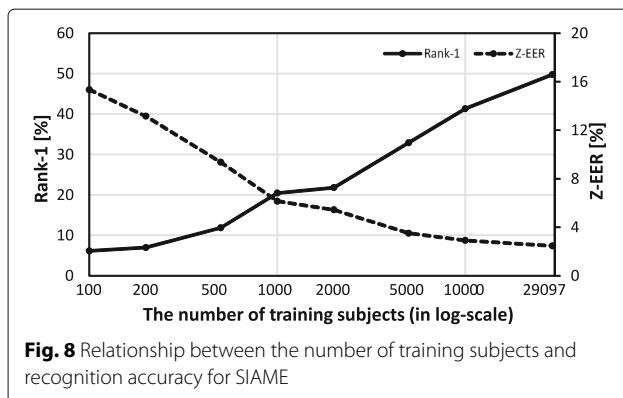


Table 4 Confusion matrix for the classification of the CS labels

a SIAME								
Actual label	Predicted label							
		No	Sb	Sm	Fr	Ba	Mu	Cp
	No	76.8	11.1	5.7	3.1	1.6	0.5	1.1
	Sb	13.5	73.0	6.1	2.5	0.9	2.1	2.1
	Sm	11.1	11.8	32.7	13.5	12.4	10.2	8.3
	Fr	3.2	2.6	7.4	72.8	0.3	4.2	9.5
	Ba	2.4	0.8	4.5	0.2	78.9	12.1	1.0
	Mu	0.9	3.2	5.9	7.6	15.9	62.6	3.9
	Cp	2.9	6.6	6.2	20.3	4.4	13.3	46.2

b mSVM								
Actual label	Predicted label							
		No	Sb	Sm	Fr	Ba	Mu	Cp
	No	72.6	9.4	11.5	2.3	1.8	0.6	1.8
	Sb	24.5	58.4	11.0	1.6	1.2	0.9	2.4
	Sm	18.1	11.7	36.0	9.8	12.2	4.2	8.0
	Fr	10.7	2.4	9.7	60.2	0.5	1.7	14.7
	Ba	5.3	1.3	6.8	0.2	74.7	10.6	1.0
	Mu	3.3	4.3	10.2	7.9	27.3	40.1	6.9
	Cp	8.2	7.0	11.8	30.4	7.9	8.2	26.5

No: NoCO; Sb: SbCO; Sm: SmCO; Ba: BaCO; Mu: MuCO; Fr: FrCO; Cp: CpCO.

classifiers for K classes), one for each pair of classes, and finally identified a class based on majority voting. By contrast, the remaining benchmark had a similar tendency to the cooperative and uncooperative settings, such as PCA_LDA, and GERF achieved nearly equal accuracy.

As for the classification accuracy of each label, NoCO and BaCO worked well because there was no CO in NoCO, and the shape and position of the CO were stable in BaCO. For SIAME, the CCRs were 76.8 and 78.9% for NoCO and BaCO, respectively, as shown in Table 4. For the case of SbCO and FrCO, the position and shape of the COs were fairly distinguished for other labels, and therefore, the classification accuracy of these labels was reasonable and nearly equal. However, SbCO was slightly confused with NoCO because of the shape similarity with respect to the upper part of the GEIs. As a result, sometimes samples of SbCO were misclassified as NoCO (see Table 4).

For the cases of SmCO, MuCO, and CpCO, the GEI features were not stable, and as a result, sometimes samples of these labels were misclassified as other labels. Because of the occlusion of COs with the subject's body for SmCO, the GEI feature was confused with that of SbCO, NoCO, and BaCO, depending on the part of the COs that was occluded, as shown in Fig. 3, and thus, samples were misclassified as SbCO, NoCO, and BaCO (see Table 4). Similarly, for the case of MuCO, it was confused with BaCO, because, as already discussed in Section 4.5, subjects typically carried, for example, a backpack in the back region together with a small object in other regions in MuCO, as shown in Fig. 3. Additionally, for the case of CpCO, subjects usually changed the CO's position from one region to another region through the front using the hands. Therefore, the GEI feature of CpCO was slightly confused with that of FrCO.

5 Conclusion and future work

In this paper, we presented a gait database that consisted of an extremely large population with unconstrained types and positions of COs, and presented a performance evaluation for vision-based gait recognition methods. This database had the following advantages over existing gait databases: (1) the number of subjects was 62,528, which was more than 15 times greater than the largest existing database for gait recognition; and (2) the CO positions were manually annotated, and gait samples were classified as seven distinct CS labels.

Together with the database, we also conducted four performance evaluation experiments. The results provided several insights, such as estimating the difficulty level among annotated CS labels based on recognition performance and the classification accuracy for CS labels.

Further analysis of gait recognition performance and the classification of the CS labels using our database is still needed. For example, we can evaluate the performance using more sophisticated and powerful DL-based approaches to gait recognition, which typically require an extremely large number of training samples but achieve state-of-the-art performance.

Endnote

¹The proposed database is an extension of the large-scale dataset that was introduced in [22].

Acknowledgements

We thank Maxine Garcia, PhD, from Edanz Group (www.edanzediting.com/ac), for editing a draft of this manuscript.

Funding

This work was partly supported by JSPS Grants-in-Aid for Scientific Research (A) JP15H01693, "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society"; Strategic Funds for the Promotion of Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government; and the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

Availability of data and materials

The database and evaluation protocol settings is available at <http://www.am.sanken.osaka-u.ac.jp/BiometricDB/GaitLPBag.html>.

Authors' contributions

MZU evaluated all the experiments and wrote the initial draft of the manuscript. MZU and TTN analyzed and discussed the evaluated accuracy. TTN and YM revised the manuscript. NT generated the GEI features. XL participated to evaluate one benchmark. YM and DM designed the study. YY supervised the work and provided technical support. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The Institute of Scientific and Industrial Research, Osaka University, Osaka 567-0047, Japan. ²The Institute for Datability Science, Osaka University, Osaka

567-0047, Japan. ³School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.

Received: 20 February 2018 Accepted: 2 April 2018

Published online: 30 May 2018

References

- Bashir K, Xiang T, Gong S (2009) Gait recognition using gait entropy image. In: Proc. of the 3rd Int. Conf. on Imaging for Crime Detection and Prevention. IET, London, pp 1–6
- Bashir K, Xiang T, Gong S (2010) Gait recognition without subject cooperation. *Pattern Recognit Lett* 31(13):2052–2060
- Bouchrika I, Nixon M (2008) Exploratory factor analysis of gait recognition. In: Proc. of the 8th IEEE Int. Conf. on Automatic Face and Gesture Recognition. IEEE, Amsterdam, pp 1–6
- Bouchrika I, Goffredo M, Carter J, Nixon M (2011) On using gait in forensic biometrics. *J Forensic Sci* 56(4):882–88
- Brian DeCann AR (2010) Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. vol 7667. SPIE, Orlando
- Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):27:1–27:27
- Chapelle O, Keerthi SS (2010) Efficient algorithms for ranking with SVMs. *Inf Retr* 13(3):201–215
- Chopra S, Hadsell R, LeCun Y (2005) Learning a similarity metric discriminatively, with application to face verification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol 1. IEEE, San Diego, pp 539–546
- Domingos P (2012) A few useful things to know about machine learning. *Commun ACM* 55(10):78–87
- Gross R, Shi J (2001) The CMU Motion of Body (MoBo) Database. Tech. rep., Carnegie Mellon University
- Han J, Bhanu B (2006) Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell* 28(2):316–322
- Hofmann M, Sural S, Rigoll G (2011) Gait recognition in the presence of occlusion: a new dataset and baseline algorithms. In: Proc. of the Int. Conf. on Computer Graphics, Visualization and Computer Vision, Plzen, pp 99–104
- Hofmann M, Geiger J, Bachmann S, Schuller B, Rigoll G (2014) The TUM Gait from Audio, Image and Depth (GAID) Database: multimodal recognition of subjects and traits. *J Vis Comun Image Represent* 25(1):195–206
- Hongye X, Zhuoya H (2015) Gait recognition based on gait energy image and linear discriminant analysis. In: Proc. of the IEEE Int. Conf. on Signal Processing, Communications and Computing (ICSPCC). IEEE, Ningbo, pp 1–4
- Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The OU-ISIR Gait Database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Inf Forensics Secur* 7(5):1511–1521
- Iwama H, Muramatsu D, Makihara Y, Yagi Y (2013) Gait verification system for criminal investigation. *IPSJ Trans Comput Vis Appl* 5:163–175
- Lee M, Roan M, Smith B, Lockhart TE (2009) Gait analysis to classify external load conditions using discriminant analysis. *Hum Mov Sci* 28(2):226–235
- Li X, Makihara Y, Xu C, Muramatsu D, Yagi Y, Ren M (2016) Gait energy response function for clothing-invariant gait recognition. In: Proc. of the 13th Asian Conf. on Computer Vision (ACCV 2016). Springer, Taipei, pp 257–272
- Lynnerup N, Larsen PK (2014) Gait as evidence. *IET Biometrics* 3(2):47–54
- Makihara Y, Mannami H, Yagi Y (2010) Gait analysis of gender and age using a large-scale multi-view gait database. In: Proc. of the 10th Asian Conf. on Computer Vision. Springer, Queenstown, pp 975–986
- Makihara Y, Kimura T, Okura F, Mitsugami I, Niwa M, Aoki C, Suzuki A, Muramatsu D, Yagi Y (2016) Gait collector: an automatic gait data collection system in conjunction with an experience-based long-run exhibition. In: Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016). IEEE, Halmstad, pp 1–8
- Makihara Y, Suzuki A, Muramatsu D, Li X, Yagi Y (2017) Joint intensity and spatial metric learning for robust gait recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, Honolulu, pp 6786–6796
- Martín-Félez R, Xiang T (2012) Gait recognition by ranking. In: Fitzgibbon AW, Lazebnik S, Perona P, Sato Y, Schmid C (eds). ECCV (1), Lecture Notes in Computer Science, vol 7572. Springer, Berlin, pp 328–341
- Martín-Félez R, Xiang T (2014) Uncooperative gait recognition by learning to rank. *Pattern Recognit* 47(12):3793–3806
- Mori A, Makihara Y, Yagi Y (2010) Gait recognition using period-based phase synchronization for low frame-rate videos. In: Proc. of the 20th Int. Conf. on Pattern Recognition. IEEE, Istanbul, pp 2194–2197
- Muramatsu D, Shiraishi A, Makihara Y, Uddin M, Yagi Y (2015) Gait-based person recognition using arbitrary view transformation model. *IEEE Trans Image Process* 24(1):140–154
- Nixon M, Carter J, Shutler J, Grant M (2001) Experimental plan for automatic gait recognition. Tech. rep., Southampton
- Nixon, MS, Tan TN, Chellappa R (2005) Human identification based on gait. Int. Series on Biometrics. Springer-Verlag, Boston
- Otsu N (1982) Optimal linear and nonlinear solutions for least-square discriminant feature extraction. In: Proc. of the 6th Int. Conf. on Pattern Recognition. IEEE, Munich, pp 557–560
- Sarkar S, Phillips J, Liu Z, Vega I, Ther PG, Bowyer K (2005) The HumanID gait challenge problem: data sets, performance, and analysis. *IEEE Trans Pattern Recog Mach Intell* 27(2):162–177
- Schultz C (2006) Digital keying methods. University of Bremen Center for Computing Technologies. Tzi 4(2):3
- Shiraga K, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2016) GEINet: View-invariant gait recognition using a convolutional neural network. In: Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016). IEEE, Halmstad, pp 1–8
- Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2017) On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Trans Circ Syst Video Technol* 28(99):1–1
- Tan D, Huang K, Yu S, Tan T (2006) Efficient night gait recognition based on template matching. In: Proc. of the 18th Int. Conf. on Pattern Recognition, vol 3. IEEE, Hong Kong, pp 1000–1003
- Tao D, Li X, Wu X, Maybank S (2006) Human carrying status in visual surveillance. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol 2. IEEE, New York, pp 1670–1677
- Uddin M, Muramatsu D, Kimura T, Makihara Y, Yagi Y (2017) MultiQ: Single sensor-based multi-quality multi-modal large-scale biometric score database and its performance evaluation. *IPSJ Trans Comput Vis Appl* 9(18):1–25
- UK Court (2008) How biometrics could change security. http://news.bbc.co.uk/2/hi/programmes/click_online/7702065.stm
- Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans Pattern Anal Mach Intell* 39(2):209–226
- Xu C, Makihara Y, Ogi G, Li X, Yagi Y, Lu J (2017) The OU-ISIR Gait Database comprising the large population dataset with Age and performance evaluation of age estimation. *IPSJ Trans Comput Vis Appl* 9(1):24
- Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Proc. of the 18th Int. Conf. on Pattern Recognition, vol 4. IEEE, Hong Kong, pp 441–444
- Zhang C, Liu W, Ma H, Fu H (2016) Siamese neural network based gait recognition for human identification. In: Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Shanghai, pp 2832–2836