Enhancing the robustness of forensic gait analysis against near-distance viewing direction differences

Daisuke Imoto1 · Manato Hirabayashi1 · Masakatsu Honma1 · Kenji Kurosawa2

Received: 12 March 2021 / Revised: 31 May 2021 / Accepted: 21 February 2022 / Published online: 26 March 2022 © The Author(s) 2022

Abstract

Gait analysis is a promising biometric technology to visually and quantitatively analyze an individual’s walking style. In Japan, silhouette-based quantitative gait analyses have been implemented as a forensic tool; however, several challenges remain owing the narrow range of application. One of the yet-unsolved issues pertains to the existence of a ‘slight’ but critical viewing direction difference, which leads to the incorrect judgment in the analyses of a person even when using deep learning-based feature extraction. To alleviate the critical viewing direction difference problem, we developed a novel gait analysis technique involving three components: 3D calibration, gait energy image space registration, and regression of the distance vector. Results of the GUI development and mock appraisal tests indicated that the proposed method can help achieve practical improvements in the forensic science domain.

Keywords 3D camera parameter calibration · Forensics · Gait energy image · Silhouette-based gait analysis · Practicality

1 Introduction

Gait analysis is a relatively novel biometric technology to visually and quantitatively analyze the information regarding the walking style and appearance of an individual [24]. This analysis has been widely adopted in the field of rehabilitation [2], and its application has been extended to the forensic science domain [4, 14, 16, 17]. In recent years, with the development of large-scale databases and deep learning technologies, many studies on gait authentication technologies have been conducted [1, 18–20, 26, 27]. However, in the presence of a viewing direction difference, the accuracy of gait analysis considerably decreases [21]. Even a ‘slight’ viewing direction difference may lead to incorrect analyses; however this aspect has not been effectively addressed.
Fig. 1  Definition of terms about viewing direction. Relationship between the terms ‘shooting direction’, ‘camera distance’, and ‘viewing direction’ and pedestrian and camera positions were defined. More precisely, viewing direction includes installation angle (optical axis direction) and focal length of the camera. The optical axis directed toward the origin in the ‘shooting direction’ situation. In other words, Fig. 1 simplifies the 6-dimensional pose model of external parameters (pose = position + orientation) to the 3-dimensional pose model which comprises 2-dimensional orientations and 1-dimensional position. In this context, viewing direction adequately considers all the 3-dimensional pose, whereas shooting direction only considers 2-dimensional orientations.

In the field of forensic science, gait analysis can be performed to identify whether two pedestrian CCTV footages (criminal footage and control footage) pertain to the same individual. The control footage may be recorded by another security camera or experimentally obtained at the site corresponding to the criminal footage or another site. In general, the forensic science tools must be accurate, objective, and explainable with a clear scope of application, among other factors. Observation-based methods have been implemented in several European countries since the 2000s [4, 16, 17]; however, it has been reported that these methods lack objectivity owing to their low quantitativeness [7]. Considering the correlation between quantitativeness and objectivity, in Japan, a silhouette-based method [14] that exploits silhouette information has been implemented to facilitate the scientific appraisal by experts (defined as the conventional method, SI Fig. 3a). This approach can account for slight differences in the clothing and belongings as the likelihood of the individual being the same (later denoted as $P(S|t)$) is calculated quantitatively by analyzing only the comparable areas in gait energy image (GEI) features by excluding/masking the non-comparable areas (in terms of the masking function). However, in this approach, the two comparative footages must satisfy several requirements such as a sufficient resolution, adequate frame rate, stable walking, and correspondence in the clothing conditions, viewing directions, and walking styles. Consequently, the range of application of this approach is quite narrow\(^1\). In particular, the influence of the viewing direction difference on the identification accuracy is significant. A critical problem of the conventional method is that it is based on the assumption that the pedestrian is sufficiently far from the camera. Specifically, as shown in Fig. 1, although the viewing direction is originally a composite of the shooting direction and camera distance (distance between the pedestrian and camera), it is considered in the conventional method that the camera distance can be neglected, and the viewing direction can be approximated to a specific shooting direction.

To clarify the significance of changes in the silhouette with the camera distance, Fig. 2a shows rendered silhouette images under two camera distance conditions: far and near\(^2\). It

---

\(^1\) As differences in viewing directions, clothes, and belongings deteriorate the analysis accuracy of an individual, the possibility of different individuals is extremely challenging to define even if the $P(S|t)$ is small.

\(^2\) The FOV value is defined as the maximum angle between the person and camera.
Fig. 2 Silhouette changes in the case of a small camera distance and slight viewing direction difference. Such changes cause problems in practical forensic gait analysis. (a) Rendering result of the 3D human body shape data at the same shooting angle for a small distance (perspective projection, FOV = 90°) and considerably large distance (orthogonal) between the camera and pedestrian. The rendering results were created using data in the KY 4D Gait Dataset [15] and the Meshlab software [6]. (b) Silhouette changes in the case of slight viewing direction differences; the angle differences in the walking courses are 0°, 5.5°, and 11° in a real space. Silhouette boundaries are indicated as red color.

can be noted that the silhouette changes under different camera distances (two vertical conditions in Fig. 2a) are extremely large. This problem is termed as projective distortion. Even in the same shooting direction, a difference in the camera distances leads to considerable silhouette changes. Examples of silhouette changes of not-far pedestrian images under a “slight” viewing direction difference are shown in Fig. 2b. For a small angle difference of the walking courses (5.5°) (Fig. 2b center), the two silhouettes in a similar gait period are notably different. Under a difference of 11° (Fig. 2b right), the two differences in the silhouettes further increases. In this scenario, if the viewing direction difference is neglected, as in the conventional method, the false rejection rate (FRR) of the same-person comparison becomes high.

In this regard, it is necessary to take into account the viewing direction difference considering the camera distance. Our preliminary analysis based on a consultation of scientific
appraisal also suggested that it is necessary to resolve the problems caused by the viewing direction difference (SI 2.2). A fundamental solution would be using a well-known 3D camera calibration technique. In a previous study [23], to solve the shooting direction difference problem, 3D camera calibration, 4D gait data \(^3\), and a view transformation model (VTM) [22] were adopted. In [23], considering that the viewing directions of two footages can be reproduced, two sets of 3D camera parameters were calibrated based on a well-known calibration technique, and the silhouette images under each condition were rendered from the 4D gait dataset by using the calibration results to assess the criteria to judge the same person. The difference in the shooting direction was addressed using VTM. However, [23] evaluated their approach using only footage with reasonably large camera distances or data with nearly the same quality silhouettes as in the training data (similar to the same domain data in Fig. 3b). In other words, practical evaluation involving different silhouette qualities and adequate tool development to ensure practical use have not been adequately conducted.

Considering these aspects, in this study, we establish a quantitative analysis method that is effective even in the presence of a slight viewing direction difference under a small camera distance. The objective is to enhance the practicality of the relevant detection approaches in the forensic science field. The main contributions of this study can be summarized as follows.

- The presence of viewing direction differences leads to critical accuracy deteriorations in the conventional method (Method I) and deep learning-based method (Method II).
- We developed a gait analysis method that contains three main components 1) re-training of the judgment criteria from 4D gait data based on 3D calibration, 2) correction of the GEI misalignment through the planar projection-based geometric view transformation model (PP-GVTM), and 3) support vector regression (RBF kernel) of the distance vector. Theoretical and practical evaluations of the proposed method (Method V) and four existing methods were performed.
- We quantitatively evaluated the relationship between the degrees of improvement of the proposed method and those of projective distortion (in terms of the camera distance and FOV).
- To enhance the practicality, we created a GUI that could execute the proposed method to facilitate the gait analysis by experts. Furthermore, mock appraisal tests containing slight viewing direction/clothing differences and notable viewing direction differences were conducted to demonstrate the effectiveness of the proposed method.

2 Methods

2.1 Gait analysis methods

Table 1 lists the analysis methods examined in this study. Methods I and Method II pertain to the conventional method and deep-learning-based method (GEINet), respectively. Method III is based on 3D calibration only, and Methods IV and V (proposed method) are based on 3D calibration in combination with GEI registration and/or distance vector regression.

\(^3\)4D refers to 3-dimensional space + time.
Fig. 3 Proposed method based on 3D camera calibration, planar projection-based GEI registration, and distance vector regression. (a) Process flow of Method V. The method only with 3D calibration (Method III) excludes both the regression and registration processes. (b) Definition of the same/different domain data to illustrate the relationship between the raw silhouette and corresponding simulated silhouette. The training data were based on the simulated silhouette data.
### Table 1 Methods compared in this study

| Index | Method’s outline | Reference |
|-------|------------------|-----------|
| I     | Conventional     | [14]      |
| II    | GEINet-based features | [26] |
| III   | 3D Calibration only (w/o VTM in [23]) | [23] |
| IV    | 3D Calibration + VTM [22] (w VTM or AVTM in [23]) | [23] |
| V     | 3D Calibration + PP-GVTM + SVR (RBF kernel) | This study (Proposed) |

### 2.1.1 Method I: Conventional method

In Method I (SI 2.2, SI Fig. 3), the most similar shooting direction pertaining to the given discrete conditions of each footage is selected. In this study, the settings selected for Method I are the same as those involving a slight shooting direction difference. Based on the pre-trained distributions of distances $t$, $P(S|t)$, the likelihood of the same person was calculated.\(^4\) Method I involves two types of error: approximation error and the error caused by ignoring the viewing direction difference. To highlight the importance of considering the viewing direction difference by excluding the heuristic approximation error, the evaluation of Method I was conducted considering that 3D calibration was performed for only one viewing direction, except in the appraisal tests, as described in 3.5.2. In the appraisal tests, flowing [14], the closest condition from 24 shooting direction conditions involving the combinations of two installation heights (80 cm and 240 cm) and twelve horizontal shooting directions in 30° increments, was selected for each footage.

### 2.1.2 Method II: GEINet

Method II (SI 2.3) is a deep-learning-based method known as GEINet, trained using approximately 1.07 million GEIs created from OU-MVLP datasets [27] involving eleven shooting directions of 10307 people. The evaluation of Method II was performed only when the whole body could be analyzed because this approach does not involve the masking function.

### 2.1.3 Three methods based on 3D calibration

Figure 3a shows the outline of the proposed method, i.e. Method V. Method V involves three main procedures: 3D camera parameter calibration (henceforth described as 3D calibration, for conciseness), GEI space registration, and distance vector regression. In 3D calibration, the internal and external parameters of two cameras of both criminal and control footages are calibrated (SI 1.1). In this case, 3D calibration involves identifying the internal parameters of the camera (lens distortion, focal length, and optical center etc.) and the position and direction of the camera in 3-dimensional space with respect to the pedestrian walking course (denoted as 3D calibration in Fig. 3a). Next, by performing rendering (or perspective projection and simulation) from a 4D gait database [15] using the calibrated viewing direction parameters, the silhouette data of each viewing direction are obtained for training (defined as silhouette videos from 4D gait data in Fig. 3a and rendering in Fig. 3b). Although the

---

\(^4\)The calculation of $P(S|t)$ was based on [13] (SI 1.2)
3D calibration-based gait analysis technique has been proposed [23], in this work, a concrete definition of the on-site procedure at the CCTV installation location is established (as described in 2.4). This method is the basis of the two methods (Methods IV and V) and named Method III.

Nevertheless, when only 3D calibration is implemented (Method III), misalignment in the GEI image space owing to the viewing direction difference remains. In the registration process, the viewpoint change in the GEI space is addressed. Muramatsu et al. [23] adopted the registration method [22] to estimate the latent features of two shooting directions through singular value decomposition (SVD). This method is defined as Method IV. As described later, Method IV is impractical because the performance is rather low when the evaluation data correspond to domains different from those pertaining to the training data. Therefore, we devised a framework for GEI space registration while maintaining the geometric constraints based on the planar projection (SI 1.3), with reference to a previous approach known as the GVTM [8]. Our registration method is denoted as PP-GVTM (planar projection-based geometric view transformation model). Although the loss value might be fixed to the local minimum due to its iterativeness, the convergence of the PP-GVTM registration algorithm was experimentally confirmed (SI 3.3.6 and SI Fig. 11). Nevertheless, even when 3D calibration and registration are applied, the GEI misalignment may not be able to be completely corrected especially under small camera distances. In this situation, if the distance between two footages is evaluated through a simple sum of each GEI, as in SI equation (13), regions with a large misalignment may be emphasized, resulting in a loss of individuality. Therefore, we performed a regression of the distance vector to obtain the distance value \( t \) as explained in SI 1.4. A comparison of the nine regression methods listed in SI Table 1, under an angle difference 11°, indicated that the RBF kernel-based support vector regression (SVR) stably exhibited a high accuracy (SI Fig. 2b). Therefore, the SVR (RBF kernel) was selected as the regression method. The method using all the procedures (3D calibration, PP-GVTM registration, and SVR (RBF kernel)) is denoted as Method V (proposed method). The proposed method was also subjected to an ablation study, excluding either PP-GVTM registration and SVR (RBF kernel), as described in SI 3.2, to demonstrate the effectiveness of using both the registration and regression processes.

### 2.2 Practical evaluation: domain definition

The two types of evaluations were conducted using the same domain data and different domain data. In the training phase, the recognition criteria were trained using the simulated silhouette images (simulated silhouette in Fig. 3b left) rendered from 4D gait data based on 3D calibration. In the test phase, practical analyses were conducted using raw silhouette images (raw silhouette in Fig. 3b right) extracted from footage videos\(^6\). These data are referred to as different domain data as the silhouette quality of the simulated and raw silhouettes may differ. The phenomenon occurs because 4D gait data are estimated data with a centimeter-order resolution in 3D space. In general, evaluation using different domain data (practical evaluation) is critical to enhance the practicality of an approach to obtain legal evidence in forensic science, which is a novelty of this research.

\(^5\)Here, the regression was implemented using the python library scikit-learn [25].

\(^6\)The silhouette extraction process was manually performed, involving manual corrections on automatically determined silhouette images by U-Net [9]. Because footages often correspond to low image qualities, it is difficult to fully automate the silhouette extraction process even when using semantic segmentation technologies.
2.3 Experimental setting

The conditions of cameras installation in a laboratory experimental room are shown in Fig. 4 and SI Fig. 7a. We set ten cameras, denoted as C1 – C10, and established three walking courses, with angle difference in real space corresponding to 0°, 5.5°, and 11°. For conditions involving C1 – C6, in addition to the evaluation using the same domain data, evaluation using different domain data was also performed. For conditions involving C7 – C10, only evaluation using the same domain data was performed. C1 – C6 could be interpreted as the combination of three horizontal shooting directions: front (C1 and C2), diagonal (C3 and C4), and lateral (C5 and C6), and two installation heights: as high as pedestrian height (C1, C3, C5) and higher positions (C2, C4, C6). It should be noted that the problem considered in this study pertains to the comparative analysis of different viewing directions when the differences in the walking course in a real space are 0°, 5.5°, and 11°. In the evaluation using the same domain data, experimental data were used to calibrate the 3D camera parameters for each walking course for each camera. In the evaluation using different domain data, eleven pedestrians were requested to walk as usual on the three walking courses for two trials, and the silhouettes were extracted from the footages on the six camera conditions (C1 – C6). The total number of footages was 376 (11 people × 2 trials × 3 walking courses × 6 cameras), and the total number of images was 16929. Finally, all the footages used for the appraisal tests, as described in 3.5.2, were additionally recorded to mimic the actual environments, and the 3D calibrations of each camera were performed. The total number of images used for the appraisal tests was 1257.

| Small Zenith Angle (β<25°) | Large Zenith Angle (β>25°) |
|----------------------------|----------------------------|
| **Small Azimuth Angle** (α<16°) | C1 | 0° | 5.5° | 11° |
| | C2 | 0° | 5.5° | 11° |
| **Medium Azimuth Angle** (16°<α<60°) | C3 | 0° | 5.5° | 11° |
| | C4 | 0° | 5.5° | 11° |
| **Large Azimuth Angle** (70°<α) | C5 | 0° | 5.5° | 11° |
| | C6 | 0° | 5.5° | 11° |

Fig. 4 Experimental settings. Examples of a representative pedestrian image in each camera condition (C1–C6). Definition of α and β are described in SI 2.4
2.4 Concrete Method for On-site 3D Camera Calibration

Although the 3D calibration method is fairly standardized (SI 1.1), to enhance the practicality, we established an on-site concrete procedure to calibrate the 3D camera parameters at the CCTV installation location. First, the 3D positions and directions of the pedestrian and camera were determined through the scene recreation of each CCTV footage. Next, after setting the coordinate system (XYZ in SI Fig. 1b) along the walking direction (Z-axis), the 2D-3D correspondences were determined by placing the 3D structure (box) at the known 3D positions. We created a sheet with a printed black and white square pattern with 50 cm squares to streamline the procedure (Fig. 5a). From the obtained 2D-3D correspondences, the projection matrix $J$ of SI (1) could be calculated by singularly decomposing the matrix $M$ of SI (3).

Figure 5b shows the calibration results of the 3D camera parameters in the laboratory (2.3). The square with yellow-blue gradation indicates the floor, and the centrally-positioned pedestrian with yellow-blue gradation is the 3D model of walking near the origin $(0,0,0)$. As mentioned previously, the viewing directions of the twelve conditions were well reproduced.

2.5 Evaluation

As the gait feature, GEI [12] for one second was considered, as described in SI 1.2. The KY 4D Gait Database (Straight) [15] and the experimental data explained in 2.3 were used for the evaluation. The 4D gait data contained a set of time series of 3D volume data created using the visual hull method, with walking data of four trials for each person recorded for 42 subjects. As mentioned in 2.3, two types of evaluations using the same and different domain data were conducted. First, in the evaluation using the same domain data, cross-validation was performed by preparing four types of data divisions, with the data of 32 and 10 participants involved in training and testing, respectively. The average and standard deviation values of the accuracy were calculated for the same person, different people, and average$^7$. For each cross-validation, the training data contained 192 and 7936 comparisons for the same person and different people, respectively, and the evaluation data contained 60 and 720 comparisons for the same person and different people, respectively. Next, in the evaluation using different domain data, based on the judgment criteria assessed from the 4D gait data for 42 participants, the accuracies were calculated considering the silhouette images extracted from the experimental footages. The training data contained 252 and 13776 comparisons for the same person and different people, respectively, and the evaluation data contained 11 and 220 comparisons for the same person and different people, respectively, for each condition. The footages analyzed in this study were set to contain approximately one – two gait periods (SI Fig. 10) considering practical analysis.

3 Results

3.1 Conventional methods in the case with viewing direction differences

First, the influence of silhouette changes under slight viewing direction differences (Fig. 2) on the analysis accuracies of the Method I was examined. The average accuracies (black

$^7$In the average condition, the ratio of the same person and different people is 1: 1.
Fig. 5  Method to achieve 2D-3D correspondence for on-site 3D camera calibration with respect to the pedestrian walking course. A black and white patterned sheet is prepared, and a 3D box of known size is placed on the sheet.

bars in Fig. 8) of Method I in the five conditions (C1 – C4, and C6) decreased significantly as the walking course angles (angle difference) increased. In the case with the angle difference of 11°, although the accuracies for different people (black bars in Fig. 6 right) were sufficiently high, the accuracies for the same person (black bars in Fig. 6 left) were zero. Specifically, the non-linear person shape (silhouette) changes on the image plane caused by projective distortion (Fig. 2) rendered it difficult to correctly analyze the same person when
the viewing direction difference was slight. Owing to the non-consideration of the viewing direction differences, the FRR for the same person comparison was high, resulting in an incorrect judgment of the same person.

The reason behind the aforementioned problem in Method I was analyzed in depth, as shown in Fig. 7. There was a tendency that the distributions of $t$ of the same person and
Fig. 7  Difference between an angle difference of 0° and 11°. Best viewed in color. (a-f) Differences of the distributions of the distance value $t$ between an angle difference of 0° and 11° in the cases of C1 – C6 are shown in (a) – (f), respectively. The light-blue, magenta, and gray dotted lines indicate $P(t|S)$, $P(t|D)$, and $P(S|t)$ for the angle difference of 0°, respectively, and the blue, red, and gray solid lines correspond to $P(t|S)$, $P(t|D)$, and $P(S|t)$ for the angle difference of 11°, respectively. (g) The resulting equal-error-rate (EER) threshold values of $t$ in the cases of the three angle differences (0°, 5.5°, and 11°) and camera conditions (C1 – C6) are analyzed.
different people ($P(t|S)$ and $P(t|D)$) in the case with an angle difference of 11° became separated from those in the case with no angle difference. It should be noted that the distributions of the 11° case are unknown in the context of Method I, which are actually the results of 3D camera calibration (Method III). As shown in Fig. 7a–f, the tendency of distribution separation was remarkable in five conditions (C1 – C4 and C6), whereas the difference in the distribution between an angle difference of 0° and 11° was small in C5. As shown in Fig. 7g, the resulting threshold values of $t$ were sensitive to angle difference values in the five conditions, resulting in the large false acceptance rates in the case with an angle difference of 11° (black bars in Fig. 6) because the values $t$ from the unknown real distributions in the 11° case were judged based on the distributions/threshold of 0°. On the other hand, the threshold in C5 was insensitive to angle difference values, which resulted in relatively high accuracy in the case with an angle difference of 11°.

It may be considered that the use of deep-learning-based methods trained on large amounts of data containing various shooting directions can likely overcome this problem. Thus, we analyzed the person verification results based on Method II. The results are shown as ochre bars in Figs. 6, 8, and 9. Figure 8 shows that the average accuracies of Method II were slightly lower than those of the conventional method when the angle difference was 0°. Although the accuracy reduction caused by the angle differences was smaller in Method II than that in Method I, a notable reduction in the accuracy occurred, especially when the angle difference was 11°. Under the angle difference of 11°, although the accuracies of different people (ochre bars in Fig. 6 right) were relatively high, the accuracies of the same person (Fig. 6 left) were low. Thus, although the average accuracy for Method II was higher than Method I in the presence of a viewing direction difference, the accuracies of Method II decreased considerably under a large angle difference.

Overall, even in the case of a slight viewing direction difference (as shown in Fig. 2b) and the use of deep-learning approaches (Method II), the accuracy deterioration for the same person occurred.

### 3.2 Effectiveness of the proposed method

#### 3.2.1 Accuracy for cases with the same person and different people

Figure 6a shows the results of the evaluation using the same domain data under camera conditions C1 – C6 for the same person and different people. Figure 6b shows the corresponding results of the evaluation using different domain data. In Method III, indicated as blue, under the five conditions (C1 – C4 and C6), the accuracies of the same person in the evaluations using the same and different domain data were approximately 70 – 86% and 46 – 100%, respectively, indicating a significant and reasonable enhancement compared to those in Methods I and II, respectively. In C5, the accuracy improved by 7 and 18 percentage points, respectively. However, the accuracies of different people under the evaluation using the same and different domain data were approximately 61 – 88% and 47 – 92%, respectively, considerably smaller than those of Method I, and the maximum decrease was approximately 53 percentage points.

When using Method IV, indicated as magenta, the accuracies of the same person in the evaluations using the same and different domain data were approximately 92 – 96% and 0 – 9%, respectively, indicating a significant difference in the accuracies in the evaluations using the same and different domain data. The accuracies for different people in the evaluations using the same domain data were 87 – 91%, slightly lower than those of Method I, albeit relatively high.
Fig. 8  Relationship between average accuracy and walking angle difference. Best viewed in color. (a,b)
Evaluation results using the same and different domain data are presented in (a) and (b), respectively.
Finally, when using Method V, indicated as red, the accuracies of the same person in the evaluations using the same and different domain data were approximately 85 – 95% and 73 – 91% under the three conditions (C1 – C3), indicating a significantly improvement over those of Method I. In the two conditions (C4 and C6), the accuracies were 88 – 94% and 40 – 64%, respectively. In C5, the accuracies improved by approximately 11 and 27 percentage points, respectively. The accuracies of different people when using Method V were approximately 79 – 92% and 86 – 97%, respectively, lower than those of Method I, with the maximum decrease being as low as approximately 21 and 14 percentage points, respectively.

The average accuracies in the evaluations using the same and different domain data are shown in Fig. 9a left and Fig. 9b left, respectively. In the evaluation using the different domain data, the averaged accuracies exhibited the following order: Method V > Method
Method III > Method II > Method I > Method IV. In particular, the average accuracy increases for Method V in the evaluations using the same and different domain data were approximately 32 and 28 percentage points compared to that in Method I. Among Methods I, III, and V (SI Tables 3 and 4), the accuracy for Method III was considerably higher than that for Method I by approximately 5 – 28 and 15 – 32 percentage points under the five conditions (C1 – C4 and C6). Method V was considerably more accurate compared to Method I, by approximately 32 – 43 and 18 – 39 percentage points, respectively, under the five conditions. Furthermore, Method V exhibited considerably higher accuracies than those for Method II by approximately 12 – 38 and 6 – 42 percentage points, respectively, under all the six conditions.

In summary, Method IV was reasonably effective in the evaluating of the same domain data, although the effectiveness decreased in the case of different domain data. The remaining methods exhibited similar accuracies in the evaluation for the same domain and different domains. The averaged accuracies for the practical evaluation of C1 – C6, as summarized in SI Table 6, indicated that Method V was the most effective method.

### 3.2.2 Evaluation of viewing direction dependence

Figure 8a shows the result of the dependence of the averaged accuracy in the evaluation using the same domain data with the viewing direction difference under different camera conditions. When using Method I, a strong dependence was found under the five conditions (C1 – C4, and C6). When using Method II, although a reasonable dependence was found under the four conditions (C1 – C4), a strong dependence was found in C6. When using Method III, although a dependence was observed, the accuracies were significantly improved compared to those in Methods I and II; nevertheless, the accuracies of Methods II and III were similar in C1 and C3. In the case of Methods IV and V, the dependence was less notable than that in Methods I – III.

Figure 8b shows the corresponding result of the evaluation using different domain data. When using Method III, although a certain angle dependence was observed, the accuracies were higher than those of Methods II among the 15/18 conditions. When using Method IV, in contrast to that in the evaluation using the same domain, a strong angle dependence was noted. Moreover, the accuracies were comparable to and higher than those obtained using Method I under the 13/18 and 2/18 conditions, respectively. The accuracies obtained using Method V were comparable to and higher than those obtained using Method III under the 12/18 and 5/18 conditions, respectively. Furthermore, under the condition with an angle difference of 11°, a significant accuracy improvement was confirmed under 5/6 conditions, except in the case of C6. In other words, Method IV was not practically effective because the results of the evaluation using the same domain and different domains differed considerably in terms of the angle dependence. Nevertheless, Method V was practically effective because the angle dependence could be significantly reduced in both the evaluations.

### 3.2.3 Evaluation of the masking function

The results of evaluation using the same domain in the case of only the upper and lower bodies being analyzable are shown in Fig. 9a center and Fig. 9a right, respectively. The corresponding results of the evaluation using different domain data are shown in Fig. 9b center and Fig. 9b right. When only the upper body was analyzable, compared to Method I, the average accuracies of Method V for the evaluations using the same and different domain data were enhanced by approximately 30 and 23 percentage points on average, respectively.
When only the lower body was analyzable, compared to Method I, the average accuracies of Method V improved by approximately 25 and 19 percentage points, respectively. Compared to that in the case with the whole body analyzable, the maximum accuracy decrease in Method V when the upper body was analyzable was approximately 4 and 14 percentage points. Although the results were highly condition dependent, the average accuracies when a part of the body was analyzable were smaller than those when the whole-body was analyzable. The practical evaluation indicated that the accuracies when only the upper and lower bodies were analyzable exhibited the following orders: Method V > Method III > Method I > Method IV, and Method V > Method III > Method I > Method IV, respectively. Overall, Method V was effective even when only a part of the body was analyzed.

3.3 Correlation analysis between accuracy improvement and projective distortion (camera distance and FOV)

To clarify the reason for the effectiveness of Method V, we attempted to evaluate the conditions in which the proposed method was effective by quantifying the parameters of the projective distortion (camera distance and FOV). The influence of the projective distortion can be quantified by a small camera distance and a large FOV value. In terms of the correlation coefficient $r$, a weak correlation, reasonable correlation, and strong correlation corresponded to $0.2 < |r| \leq 0.5$, $0.5 < |r| \leq 0.8$, and $0.8 < |r|$, respectively.

The relationship between the minimum camera distances and degrees of accuracy improvements under the six conditions (C1 – C6) is shown in Fig. 10 left. A smaller camera distance corresponded to a greater degree of improvement owing to a strong correlation ($r = -0.87$). The relationship between the maximum FOV and degrees of accuracy improvements under the six conditions is shown in Fig. 10 right. A larger FOV value corresponds to a greater degree of improvement owing to a reasonable correlation ($r = 0.65$). These findings demonstrate that a higher accuracy improvement occurs under a strong projective distortion.

3.4 Practicality comparative analysis

3.4.1 GUI development

The proposed method can be divided into two modules, the training phase (3D calibration and training) and testing phase (verification). Although the verification part has been reported in the previous work [14], the work did not focus on the former part. Therefore, first, we developed the GUI of the former part to enable appraisal assistance (Fig. 11a). In general, after examining the 2D-3D correspondence data prepared in advance, 3D calibration can be performed by clicking the <Calibration> button, and the positions/directions of the two cameras with respect to a pedestrian can be clarified in the <Calibration Results> part. The calibration results of both the criminal footage (Camera 1) and control footage (Camera 2) can be illustrated (the camera with the red and blue arrows, respectively). After selecting the closest frame rate, the user can specify the directory of the training result by clicking the <Set output dir. of training> button. Subsequently, the training process can be executed by clicking the <Training w/ PP-GVTM> button.

Next, to enable analysis that reflects the trained results in the former 3D calibration and training part, the GUI of the verification part was created with reference to the aforementioned previous work [14] (Fig. 11b). In general, after inputting the footage with the corrected lens distortion and corresponding silhouette images, the GEIs can be calculated.
Correlation between improvement in the accuracy and minimum pedestrian-camera distance and maximum FOV. Correlation between the accuracy improvement and minimum camera distance is indicated on the left, and the correlation between the accuracy improvement and maximum FOV is indicated on the right. Evaluation results based on the different domain data (practical evaluation) under the angle difference of 5.5° are considered.

Mock appraisal tests under challenging practical situations

Mock appraisal tests were conducted using the abovementioned appraisal assistance software. Correctness of typical results of the same person and different people obtained using Method V in Condition I was confirmed (SI Fig. 9). Table 2 summarizes the analysis results.

| Condition   | Method I: Conventional | Method V: 3D Calibration + PP-GVTM + SVR (RBF kernel) |
|-------------|------------------------|-----------------------------------------------------|
| Condition I: Same person | 0/4 (0%) | 3/4 (75%) (+ 75 percent points) |
| Condition I: Different people | 12/12 (100%) | 9/12 (75%) (- 25 percent points) |
| Condition II: Same person | 0/3 (0%) | 3/3 (100%) (+ 100 percent points) |
| Condition II: Different people | 6/6 (100%) | 3/6 (50%) (- 50 percent points) |
| Condition III: Same person | 0/4 (0%) | 2/4 (50%) (+ 50 percent points) |
| Condition III: Different people | 8/8 (100%) | 8/8 (100%) (± 0 percent points) |
Fig. 11 GUI development for forensic use including 3D camera calibration with respect to the pedestrian walking course and 3D camera calibration based gait analysis (human verification) system. Best viewed in color: (a) GUI for 3D camera calibration and training. (b) GUI for gait analysis (human verification).
of Methods I and V under Conditions I – III (Fig. 12). When using Method I, all the analyses of different people were correct, whereas all the analyses of the same person were wrong. In contrast, when using Method V, although a somewhat reduced accuracy for the different people case occurred, the accuracies for the same person case improved significantly, resulting in notable improvements in the average accuracies.

4 Discussion and conclusion

In this study, we highlighted the problem of the reduction in the verification accuracy of the same person in Methods I and II under different viewing directions, especially under small camera distance. We developed a solution to this problem and quantitatively evaluated its practical effectiveness through evaluations involving different domain data (practical evaluation). It was noted that the proposed method considerably outperformed Methods I and II in terms of the accuracy. Furthermore, to enhance the practicality of the approach, in addition to establishing on-site concrete procedures for 3D calibration at the CCTV installation location, a set of GUIs that could execute Method V was created. The appraisal tests results demonstrated the superiority of Method V over the other methods.

Overall, Method V exhibits high forensic practicality in terms of gait analysis owing to its robustness against the dependence on the viewing direction difference. The average accuracy is enhanced significantly when using Method V as it can overcome the viewing direction problem, which cannot be effectively addressed using conventional approaches. Nevertheless, although more appropriate judgment criteria can be implemented by introducing Method V, errors in the case of different people cannot be entirely avoided, and this aspect must be considered during practical use. In particular, when making a judgment considering only one comparison in practical use, the user must verify the corresponding condition with the gait data of the other people. Furthermore, there is a drawback that it is not possible to know whether or not Method V can be effective without looking at the proposed analysis results by visiting the CCTV installation location. Thus, since the cost of visiting the installation location is often significant, if an accurate calculation-based method for 3D calibration is developed in the future, it may be possible to screen whether or not to
visit the CCTV location to perform the proposed gait analysis. Such a calculation method does contribute to the practical use.

Furthermore, as described in 3.4, a smaller camera distance of larger FOV value corresponds to a greater degree of accuracy improvement. Therefore, the accuracy decrease for the same person case cased by the viewing direction difference under a small camera distance can likely be attributed to the projective distortion. Although this correlation in the context of other image recognition problems has not been widely reported, the problem caused by projective distortion is likely significant for all such applications. Furthermore, it should be noted that “small camera distance” tested in this study (SI Table 7) is consistent with the real installation of common security cameras [5, 10] and the distance categorization defined in another biometrics study [11] (SI 3.3.7).

As mentioned previously, in the context of deep learning, the existing databases and the related analyses [20, 26, 28, 29] considered only the shooting directions, and the camera distance was assumed to be sufficiently large. In other words, the databases could not consider conditions in which the camera distance was small. Moreover, as illustrated in Fig. 2a, the silhouette shapes change significantly even in the same shooting direction under a small camera distance. Training based on those gait databases could not lead to a view-invariant scenario even if the concept of the study was based on this context (Fig. 2b). It is considered that the accuracy can be enhanced by constructing small camera distance or large FOV, although future work remains necessary.

The results demonstrate that the problem of angle difference in gait analysis can be treated quantitatively and scientifically, and these findings can help enhance the applicable range of quantitative gait analysis. In terms of observation-based gait analysis, Birch et al. stated the need to match the viewing directions [3], and Edmond et al. disputed the scientific basis of gait analysis [7]. However, we believe that our findings can contribute to the relaxation of the viewing direction conditions in practical quantitative gait analysis, thereby enhancing the reliability of gait analysis, as applied in the forensic science domain.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11042-022-12751-0.

Acknowledgements The authors would like to thank Prof. Yasushi Makihara at Osaka University for the valuable discussions. A part of this study was supported by JSPS KAKENHI grants (JP17K18379 and JP20K19835).

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

1. Andriluka M, Pishchulin L, Gehler P, Schiele B (2014) 2D human pose estimation: new benchmark and state of the art analysis. In: 2014 IEEE Conference on computer vision and pattern recognition, pp 3686–3693
2. Baker R (2006) Gait analysis methods in rehabilitation. Journal of NeuroEngineering and Rehabilitation, Vol 3, Article number: 4
3. Birch I, Raymond L, Christou A, Fernando MA, Harrison N, Paul F (2013) The identification of individuals by observational gait analysis using closed circuit television footage. Sci Justice 53(3):339–342
4. Bouchrika I, Goffredo M, Carter J, Nixon M (2011) On using gait in forensic biometrics. J Forensic Sci 56(4):882–889
5. CCTV42 Understanding the range of a CCTV camera. https://www.cctv42.co.uk/help-advice/faqs/can-you-explain-about-the-range-of-a-cctv-camera/ Accessed: 13th May 2021
6. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G (2008) Meshlab: An open-source mesh processing tool. In: Scarrano IV, Chiara RD, Erra U (eds) Eurographics Italian chapter conference. The Eurographics Association
7. Edmond G, Cunliffe E (2016) Cinderella story? the social production of a forensic “Science”. J Crim Law Criminol 106:219–273
8. El-Alfy H, Xu C, Makihara Y, Muramatsu D, Yagi Y (2017) A geometric view transformation model using Free-Form deformation for Cross-View gait recognition. In: 2017 4Th IAPR asian conference on pattern recognition (ACPR), pp 929–934
9. Falk T, Mai D, Bensch R et al (2019) U-net: Deep learning for cell counting, detection, and morphometry. Nat Methods 16:67–70
10. Farsight Security Services Ltd. Where to Position CCTV Cameras in Your Business. https://www.farsight.co.uk/blog/where-to-position-cctv-cameras-in-your-business/ Accessed 13th May 2021
11. Guo BH, Nixon MS, Carter JN (2019) Soft biometric fusion for subject recognition at a distance. IEEE Trans Biom Behav Identity Sci (T-BIOM) 1:292–301
12. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
13. Imoto D, Kurosawa K, Tsuchiya K, Kuroki K, Hirabayashi M, Akiba N, Kakuda H, Tanabe K, Hawai Y (2018) A novel human identification method by gait using dynamics of feature points and local shape features. In: 2018 7Th the european workshop on visual information processing (EUVIP), pp 1–6
14. Iwama H, Muramatsu D, Makihara Y, Yagi Y (2013) Gait verification system for criminal investigation. Inf Media Technol 8(4):1187–1199
15. Iwashita Y, Baba R, Ogawara K, Kurazume R (2010) Person Identification from Spatio-temporal 3D Gait. In: 2010 international conference on emerging security technologies, pp 30–35
16. Larsen PK, Simonsen EB, Lynnerup N (2008) Gait analysis in forensic medicine. J Forensic Sci 53(5):1149–1153
17. Larsen PK, Simonsen EB, Lynnerup N (2010) Use of photogrammetry and biomechanical gait analysis to identify individuals. In: 2010 18th European signal processing conference, pp 1660–1664
18. Li X, Makihara Y, Xu C, Yagi Y, Yu S, Ren M (2020) End-to-end model-based gait recognition. In: Proceedings of the Asian conference on computer vision (ACCV)
19. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision – ECCV 2014. Springer International Publishing, Cham, pp 740–755
20. Makihara Y, Mannami H, Tsuji A, Hossain MA, Sugira K, Mori A, Yagi Y (2012) The OU-ISIR gait database comprising the treadmill dataset. IPSJ Trans Comput Vis Appl 4:53–62
21. Makihara Y, Matosvki D, Nixon MS, Carter JN, Yagi Y (2015) Gait recognition: Databases, representations, and applications. In: Encyclopedia of electrical and electronics engineering
22. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: Leonardis A, Bischof H, Pinz A (eds) Computer vision – ECCV 2006. Springer, Berlin, pp 151–163
23. Muramatsu D, Shiraishi A, Makihara Y, Uddin MZ, Yagi Y (2015) Gait-Based Person recognition using arbitrary view transformation model. IEEE Trans Image Process 24(1):140–154
24. Nixon MS, Carter JN (2006) Automatic recognition by gait. Proc IEEE 94(11):2013–2024
25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in python. J Mach Learn Res 12:2825–2830
26. Shiraga K, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2016) GEINet: View-invariant gait recognition using a convolutional neural network. In: 2016 International conference on biometrics (ICB), pp 1–8
27. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2018) Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Trans on Comput Vis Applic 10, Article number:4
28. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18Th international conference on pattern recognition (ICPR’06), 4, pp 441–444
29. Zheng S, Zhang J, Huang K, He R, Tan T (2011) Robust view transformation model for gait recognition. In: International conference on image processing (ICIP)

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