Person Tracking Based on Gait Features from Depth Sensors

Takafumi Mori[1(✉)] and Hiroaki Kikuchi[2]

[1] Graduate School of Advanced Mathematical Sciences, Meiji University, Tokyo 164-8525, Japan
cs172059@meiji.ac.jp
[2] School of Interdisciplinary Mathematical Science, Meiji University, Tokyo 164-8525, Japan
kikn@meiji.ac.jp

Abstract. Gait information is a useful biometric because it is a user-friendly property and gait is hard to mimic exactly, even by skillful attackers. Most conventional gait authentication schemes assume cooperation by the subjects being recognized. Lack of cooperation could be an obstacle for automated tracking of users and many commercial users require new gait identification schemes that do not require the help of target users. In this work, we study a new person-tracking method based on the combination of some gait features observed from depth sensors. The features are classified into three groups: static, dynamic distances, and dynamic angles. We demonstrate with ten subjects that our proposed scheme works well and the accuracy of equal error ratio can be improved to 0.25 when the top five features are combined.

1 Introduction

With the popularization of the Internet, the necessity of safe personal authentication is increasing. Conventional, knowledge-based authentication, i.e., password or PIN, is unsafe because people often forget the confidential information and there is a risk of the leakage of personal information. As a result, biometric authentication, which uses individual biological attributes, is becoming popular. In this study, we focus on a method for gait authentication that uses features of a subject's style of walking.

Gait authentication has three basic classes: machine vision [1], floor sensors, and wearable sensors [2]. A machine vision-based system authenticates people from a camera located at a distant position so that it can be used for wide-range monitoring. This allows people to be observed without being noticed. Because of these features, gait authentication is considered to be an appropriate method for individual tracking.

Authentication tests whether a given user is registered to a system, while tracking distinguishes two users appearing at distinct locations. Authentication is mainly used to prove that a target is a genuine user when the user logs in to

L. Barolli et al. (Eds.): NBiS 2018, LNDECT 22, pp. 743–751, 2019.
https://doi.org/10.1007/978-3-319-98530-5_65

a system or enters a building. However, tracking is used to identify major walking paths and to obtain statistical information about people's flow paths for marketing or crime prevention.

We found the following differences between authentication and tracking.

- People are cooperative with authentication because they are willing to use it. However, tracking is done while the target users are unconsciously.
- Tracking does not require high accuracy because the data are used for statistical information.
- Tracking should address privacy concerns because target users are not always willing to be tracked.
- Data from tracking should be removed when a person refuses to provide his or her information (opt-out).
- A threat to authentication is a malicious adversary who pretends to be the target user. A threat to tracking is to pretend to be someone else.

The differences between authentication and tracking are summarized in Table 1.

Table 1. Differences between tracking and authentication

	Tracking	Authentication
Application	Statistical information	Prove that i am a proper user
Target	Noncooperative	Cooperative
Desired accuracy	Low	High
Matching	$m : n$	$1 : n$
Privacy care	Necessary	Unnecessary
Threat	Recognizing as other person	Pretend to be proper user

In this paper, we study the use of gait information to track people. We propose a new person-tracking method using gait and demonstrate its feasibility with a trial implementation with the Microsoft Kinect V2.

Our main result is that the proposed scheme performs accurately and the equal error rate (EER) is 0.22 in the optimal case when several features are combined.

2 Related Work

A silhouette image is often used in gait authentication. For example, Han et al. proposed the gait energy image (GEI) [3], which is an average silhouette image of one cycle of walking. It requires less processing time and reduces the storage requirement and the robustness against noise.

Shiraga et al. proposed GEINet, an authentication system using GEI [1]. They used a convolutional neural network (CNN), and achieved high accuracy in image recognition when classifying GEI images. They demonstrated that gait can be used to authenticate people with high accuracy.

Muaaz et al. proposed a smartphone-based gait authentication method [2]. They used acceleration vectors observed with a smartphone in a pocket as features. A cycle of walking was used as a template. Multiple templates per subject were stored and registered. In authentication, the dynamic time warping (DTW) distance between a given cycle of walking and each template was calculated and a subject who had more than fifty percent of features for which DTW distances are less than a predetermined threshold was accepted. Moreover, they empirically proved that a mimic attack is impossible in gait authentication.

Some methods have used silhouette images. Andersson and Araujo proposed a gait authentication method using skeleton information from Kinect [4] in 2015. Igual et al. proposed a gender recognition method using depth information [5].

3 Proposed Method for Gait Recognition

In this study, we propose a gait recognition method based on sequences of three-dimensional coordinates of joints in walking. Our proposed method consists of the following five steps:

1. Data capture,
2. Cycle extraction,
3. Sequence extraction,
4. Features calculation, and
5. Identification.

3.1 Data Capture

We use Kinect V2, a motion capture device developed by Microsoft. Kinect was designed for Xbox players who control the Xbox using images of their bodies while playing. Kinect is described as a natural user interface (NUI).

Kinect facilities include an RGB camera, a depth camera, and a microphone. It identifies three-dimensional coordinates of joints of the player to recognize the player's movement. The three-dimensional coordinates captured by Kinect are called skeleton data and can be retrieved via the Kinect Software Development Kit. The specifications of the Kinect V2 are shown in Table 2.

3.2 Cycle Extraction

In this phase, we extract a cycle of walking, which is defined as a series of features in walking at which a foot reaches the same position.

To identify a cycle from continuous skeleton data, we calculate the distance between two ankles and smooth it by taking averages of two neighboring data points. Finally, we identify a cycle that begins at the first peak of distances and ends at the third peak. We show an example of ankle distance and a cycle in Fig. 1. In this figure, a cycle is from point A to point E.

Table 2. Specification of kinect V2

Attribute	Value
RGB resolution	1920 × 1080 pixel
Depth resolution	512 × 424 pixel
Frame rate	30[fps]
Num of observable people	6
Num of observable joints	6
Measurable distance	0.5–4.5 m

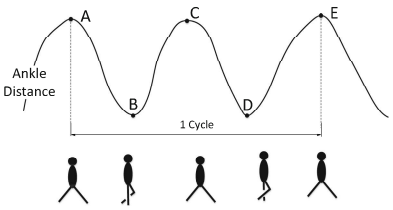

Fig. 1. Sample of one cycle

3.3 Sequence Extraction

We define a total of 36 features that are classified into three groups: static distances, dynamic distances, and joint angles.

Static distances are lengths of between adjacent joints. Because the distance between two adjacent joints is determined by the length of the bone between them, it is a stable quantity. For example, the distance S_1 is the distance between two joints in the left foot and left ankle. The static distances are illustrated with the skeleton model in Fig. 2.

Dynamic distances are distances between two arbitrary joints in a body. The distance varies with movements of the feet and arms while walking. For instance, the distance between the two feet, D_1, fluctuates periodically when a subject is walking. The dynamic distances are illustrated in Fig. 3.

Dynamic Angles are measured between the vertical and a line connecting two joints. They are illustrated in Fig. 3. Dynamic angles are also dynamic quantities.

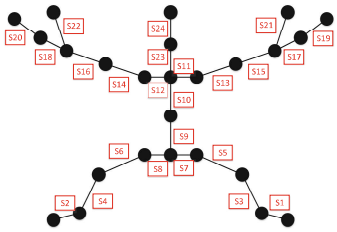

Fig. 2. Diagram of static distances

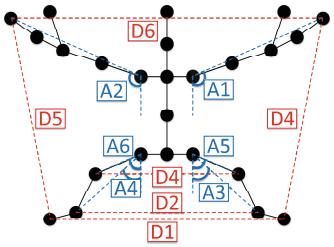

Fig. 3. Diagram of dynamic distances and dynamic angles

3.4 Statistics of Features

In this section, we calculate some useful statistics of features. The features vary in a given cycle; therefore, we use statistics of series of features in a cycle. The statistics include maximum, median, and duration of one cycle in our study.

3.5 Identification

We first consider a simple identification with a single feature and then extend it to the fusion version that combines multiple features.

Let f be a feature of walking, i.e., $f \in \{S_1, \ldots, S_{24}, D_1, \ldots, D_6, A_1, \ldots, A_6\}$. Let $f_{i,k}$ be the kth cycle of the ith subject. Statistics of a series of features $f_{i,k}$ are $\mu(f_{i,k}), median(f_{i,k}), max(f_{i,k})$.

Given two statistics (means) of the kth feature $\mu(f_{i,k})$ and $\mu(f_{j,k'})$, we say that subjects i and j are identical (the same) if:

$$same(i,j) = \begin{cases} T & \text{if } |\mu(f_{i,k}) - \mu(f_{j,k'})| \leq \theta \\ F & \text{otherwise,} \end{cases}$$

where θ is a threshold of matching. The mean can be replaced by other statistics such as the median or maximum.

Next, we extend the simpler identification by combining several features. For example, using Euclidian distance, two features f and g can be tested jointly to identify if i and j are the same:

$$same(i,j) = \begin{cases} T & \text{if } \sqrt{(\mu(f_{i,k}) - \mu(f_{j,k'}))^2 + (\mu(g_{i,k}) - \mu(g_{j,k'}))^2} \leq \theta \\ F & \text{otherwise.} \end{cases}$$

The same steps are applied to the median and maximum values. Note that the Euclidian distance can still be used when combining more than three features.

We may optimize the threshold θ to be the EER. An EER is an error rate at which $\mathrm{FAR}(\theta^*_\ell) = \mathrm{FRR}(\theta^*_\ell)$ at the optimized value θ^*_ℓ, where the false acceptance ratio (FAR) is a fraction of faulty authenticated imposter subjects and the false rejection ratio (FRR) is a fraction of genuine subjects who are wrongly judged as imposters.

4 Experiment

4.1 Experiment Purpose

The purposes of our experiment are as follows:

1. Identify efficient features that can be used to recognize a person accurately,
2. Measure how accurately persons are identified for each set of features, and
3. Find out the best method to combine features to maximize the accuracy of recognizing people.

4.2 Method

We capture ten subjects walking using Kinect V2. Each subjects walks six times. The experiment term was from 5/August/2017 to 17/August/2017. Subjects are uniquely identified with labels U_1–U_{10}.

According to Sect. 3, we set the threshold θ so that FAR is equal to FRR at θ for every feature.

4.3 Results

4.3.1 Data Capture and Calculation of Features

An example of a series of skeleton data is shown in Fig. 4, which shows two-dimensional coordinates of five typical joints, a head, both hands and both ankles, observed from the Kinect sensor over a few cycles.

4.3.2 Calculation of Features and Statistics

The statistics for $\mu(D_5)$ are shown in the bar plot of Fig. 5. Most subjects can be clearly distinguished from each other, except U_9 and U_{10}.

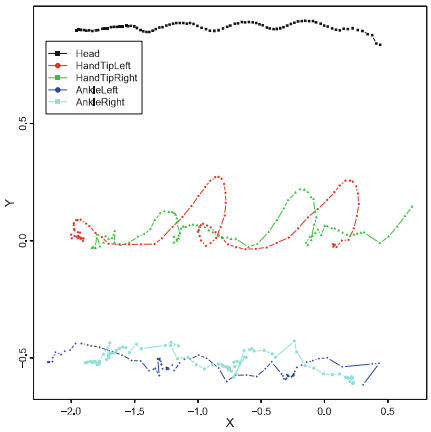

Fig. 4. Example of captured data

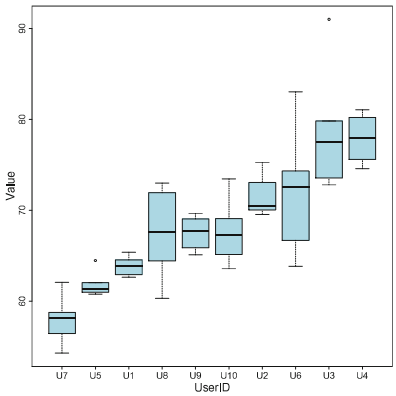

Fig. 5. Distributions of $\mu(D_5)$ for all users

4.3.3 Detection Threshold

A threshold of matching should be carefully determined by looking at the distributions of features. For example, Fig. 6 shows two histograms of $\mu(D_5)$, one for the same subject and the other for between subjects. The figure shows that the variance of distances in the same subject is smaller than that with others.

Receiver operating characteristic (ROC) curves, which show the tradeoff of FAR and FRR for the representative statistics $\mu(S_6)$, $median(D_1)$, and $max(A_2)$ are shown in Fig. 7. We found that $\mu(S_6)$ is the best feature in terms of FAR and FRR for three candidates.

4.4 Evaluation

4.4.1 Comparison of Features

A list of the top 10 EERs in ascending order is shown in Table 3.

Note that there is only one feature using max. Generally, dynamic angles are less useful in identifying subjects and only two features of dynamic angles are in the top 10 list.

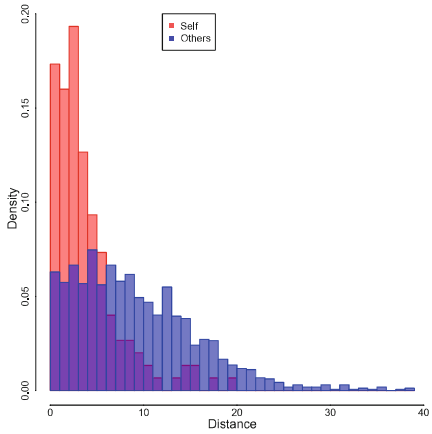

Fig. 6. Histogram of $\mu(D_5)$

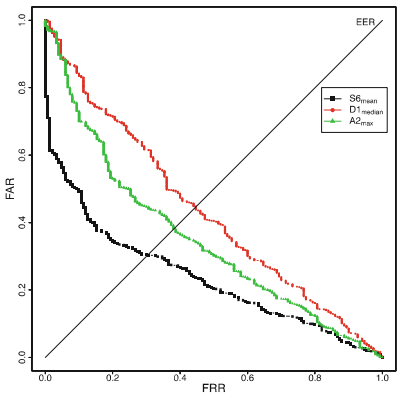

Fig. 7. ROC curves of $\mu(S_6)$, $median(D_1)$ and $max(A_2)$

Table 3. EERs arranged in ascending order

Features	Group	Statistics	EER
$\mu(D_5)$	Dynamic distances	Mean	0.29
$max(D_5)$	Dynamic distances	Max	0.29
$\mu(S_6)$	Static distances	Mean	0.30
$median(D_5)$	Dynamic distances	Median	0.30
$\mu(S_5)$	Static distances	Mean	0.31
$median(A_4)$	Dynamic angles	Median	0.31
$median(S_5)$	Static distances	Median	0.31
$median(S_6)$	Static distances	Median	0.31
$\mu(D_4)$	Dynamic distances	Mean	0.32
$\mu(A_4)$	Dynamic angles	Mean	0.32

4.4.2 Combined Features

We combined multiple features including dynamic distances and dynamic angles. The top 10 EERs listed in ascending order are shown in Table 4.

We show boxplots of EER with respect to the number of features combined using the maxima of dynamic distances in Fig. 8. The EER decreases as the number of combined feature increases, but is saturated at five and no improvement is obtained with six or more features. Combining some features helps to decrease FAR, but increases FRR. Hence, combining too many features could spoil the FAR and result in low EER. Therefore, we conclude that the best number of features to combine is around five in our experiment.

The histogram of maxima of dynamic distances as calculated for self and with others is shown in Fig. 9.

Table 4. Top 10 EERs arranged in ascending order

Features	Group	Statistics	EER
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{16})$	Static distances	Mean	0.22
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{13})$, $\mu(S_{16})$	Static distances	Mean	0.22
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$	Static Distances	Mean	0.23
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{13})$	Static distances	Mean	0.23
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{13})$, $\mu(S_{20})$	Static distances	Mean	0.23
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{13})$, $\mu(S_{16})$, $mu(S_{20})$	Static distances	Mean	0.23
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_9)$, $\mu(S_{13})$	Static distances	Mean	0.23
$\mu(S_5)$, $\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{13})$	Static distances	Mean	0.23
$\mu(S_3)$, $\mu(S_2)$, $\mu(S_9)$, $\mu(S_{16})$	Static distances	Mean	0.23
$\mu(S_5)$, $\mu(S_3)$, $\mu(S_2)$, $\mu(S_{22})$, $\mu(S_9)$, $\mu(S_{16})$	Static distances	Mean	0.23

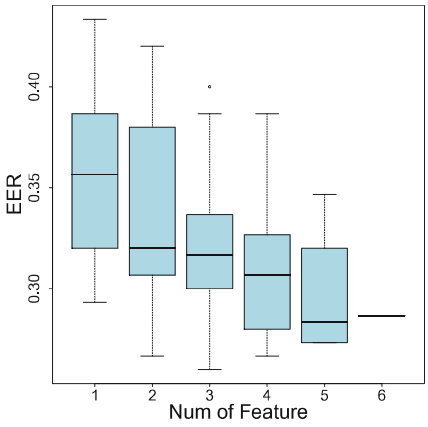

Fig. 8. Correlation of number in combination and EER

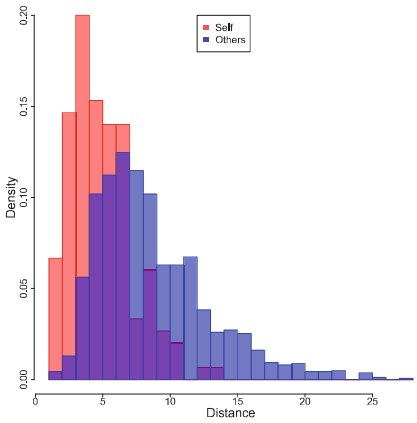

Fig. 9. Histogram of maxima of dynamic distances

4.5 Discussion

Let us consider reasons why some subjects have low accuracy. The cause of outliers in Fig. 5 is considered to be due to measurement errors of the Kinect, which tracks joints based on the image observed from a camera. Hence, if a point is hidden by the body, the Kinect cannot track the joint. Instead, the Kinect tries to estimate the coordinates of hidden joints with some estimation errors.

As for statistics, we found that maximum values are worse than mean and median, as shown in Tables 3 and 4. We think that for maxima, outliers have more significant effects than for the other statistics, which contributes to failure of identification.

We used three kinds of features, static, dynamic and angle. Based on the results in Tables 3 and 4, we found that dynamic distances perform effectively for identification. Although static distances depend on the size of the body and

dynamic angles are affected by movements of arms and feet, dynamic distances are affected by both size and movement. We believe that dynamic distances are better features.

5 Conclusions

In this paper, we have proposed a new gait recognition method and demonstrated it using a trial implementation system with Kinect V2. The best EER in our experiment was 0.22 and we conclude that the proposed method can be used for individual tracking in practical situations.

In future work, we plan to study other gait recognition methods that have lower EER or are more robust against walking noise, e.g., shoes and luggage.

References

1. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y.: GEINet: view-invariant gait recognition using a convolutional neural network. In: Proceedings of the 8th IAPR International Conference on Biometrics (ICB 2016), pp. 1–8, Halmstad, Sweden, June 2016
2. Muaaz, M., Mayrhofer, R.: Smartphone-based gait recognition: from authentication to imitation. IEEE Trans. Mob. Comput. **16**, 3209–3221 (2017)
3. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
4. Andersson, V., Araujo, R.: Person identification using anthropometric and gait data from kinect sensor. In: AAAI Conference on Artificial Intelligence (2015)
5. Igual, L., Lapedriza, À., Borràs, R.: Robust gait-based gender classification using depth cameras. EURASIP J. Image Video Process. **2013**(1), 1–11 (2013)