Gait-based Person Authentication by Wearable Cameras

Kohei Shiraga, Ngo Thanh Trung, Ikuhisa Mitsugami, Yasuhiro Mukaigawa and Yasushi Yagi
The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, 567-0047, Osaka, Japan
Email: {shiraga, trung, mitsugami, mukaigaw, yagi}@am.sanken.osaka-u.ac.jp

***Abstract*—In this paper, we propose a novel gait-based person authentication by wearable cameras, and quantitatively evaluate its authentication accuracy. In contrast to previous methods using motion sensors, we utilize wearable cameras for personal authentication by examining motion during walking obtained from visual information. This motion appears to provide a valid representation of the gait of the wearer. We then developed a prototype system of wearable surveillance, which is a new concept of surveillance we have proposed. The performance of our gait-based person authentication method was experimentally evaluated with 39 subjects using the prototype system, revealing an Equal Error Rate (EER) of 5.6%.**

I. Introduction

An increasing number of surveillance cameras have been installed in many cities around the world, in an attempt to increase safety in various places such as stations, schools, banks, shops and the entrances of private homes. However, because it is impractical to cover all areas with surveillance cameras, blind areas exist in which crimes can take place.

We have proposed a new concept of wearable surveillance (WS). In the WS, as shown in Fig. 1, a user is equipped with the WS system which is composed of some sensors, such as cameras, a computer for processing the sensory data. The system should of course be compact and lightweight enough for the user to wear and perform in a stand-alone environment without an external power source. The WS system seeks to increase the user's safety by autonomously observing his/her surroundings. This system also should be designed to be able to sense the following abnormal situations: (i) the user is deprived of the WS system by another person, (ii) a suspicious person is approaching the user, and (iii) the user is chased by a suspicious person. In addition, the WS system should be equipped with a network device so as to send alerts to a distant person, parents of the user for example. Note that in these days the system can connect to the network from almost anywhere, which means the user is always guarded by the person.

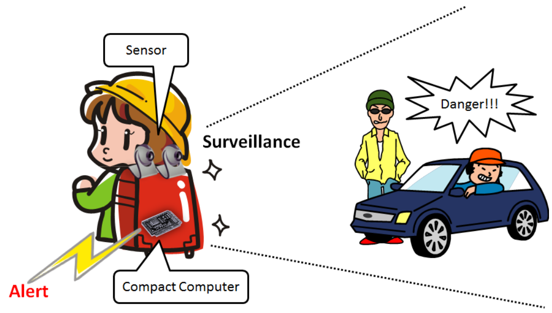


Fig. 1. A concept of wearable surveillance. The wearable surveillance system which consists of sensors, such as cameras, a compact computer and a network equipment. This system seeks to increase the user's safety by autonomously observing the ambient environment, and has capable of notifying someone in another place of information around the user.

To realize such the system, one of the important techniques is authentication of a person wearing the system. If the personal authentication is achieved, the system can judge whether that the user is wearing the system or it is deprived by another person as described in (i).

In this paper, therefore, we propose a gait-based person authentication for WS, in which gait information is acquired by wearable cameras installed on the WS system. This approach can achieve all of the functions listed above, including personal authentication using wearable cameras. Specifically, a sequence of rotation of the cameras themselves is first estimated from their images. The sequence of rotation of cameras accurately represents the gait of the wearer. Next, the sequence is segmented into distinct gait periods, and the dissimilarity between each of them and each period in registered samples is computed, then authenticate based on such dissimilarity.

A number of previous studies have examined personal

978-1-4673-1786-3/12/$31.00 ©2012 IEEE

authentication using motion sensors attached to an ankle or a backpack [1][2][3][4][5]. Their systems are suitable for measuring the motion of a pedestrian, and have been found capable of authentication with high accuracy. In contrast, our proposed method estimates gait information from images of cameras.

The images make it possible to reconstruct 3-D scene around the wearer by structure-from-motion techniques as well as estimate his/her motion. The 3-D scene information is considered to be a beneficial clue in the above mentioned cases. For example, in the case (ii), detailed positions and directions of detected suspicious individuals can be obtain by using such 3-D reconstruction techniques.

Our contribution is to reveal the possibility of the gait-based authentication using gait information obtained by wearable cameras. To evaluate our method, we developed a prototype system for the WS system designed for school children. The experimental results reveal that the accuracy of our method is approximately equal to those of previous methods using motion sensors.

In Section II, we present our personal authentication method. Section III describes our developed prototype system. Section IV shows the experimental evaluation of our method and the results of the analysis. Section V discusses the limitation of our method. Section VI summarizes the paper and describes possible future work.

II. Vision-based personal authentication

We describe our gait-based person authentication method for the WS system. The basic components of the WS system are shown in Fig. 2. The WS system is constructed based on a backpack. A stereo camera pair is installed on both sides of the top of the backpack, connected to a compact computer inside the backpack. The system is powered by a battery.

As shown in Fig. 3, first, the rotation of cameras installed on the WS system is estimated as information about the gait of the wearer. Although camera rotation is not a direct measure of the wearer's gait, it provides a valid representation of its important features. Next, the obtained gait information is segmented into distinct gait periods. Finally, the wearer is authenticated by the distance between each period and the gallery, which is previously registered period set. These processes are described in detail below.

A. Rotation estimation

Vision-based egomotion estimation has been well studied in the field of robotics and computer vision [6][7][8][9] for a number of applications. Camera egomotion generally involves six-dimensional information, consisting of three-dimensional translation and three-dimensional rotation, and it is difficult to stably estimate each motion type simultaneously. Trung et al. [10]

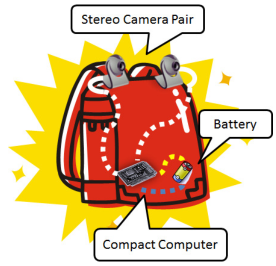


Fig. 2. Components of the WS system. A stereo camera pair is installed on both sides of the top of the backpack, and connected to a compact computer inside the backpack. This system powered by a battery.

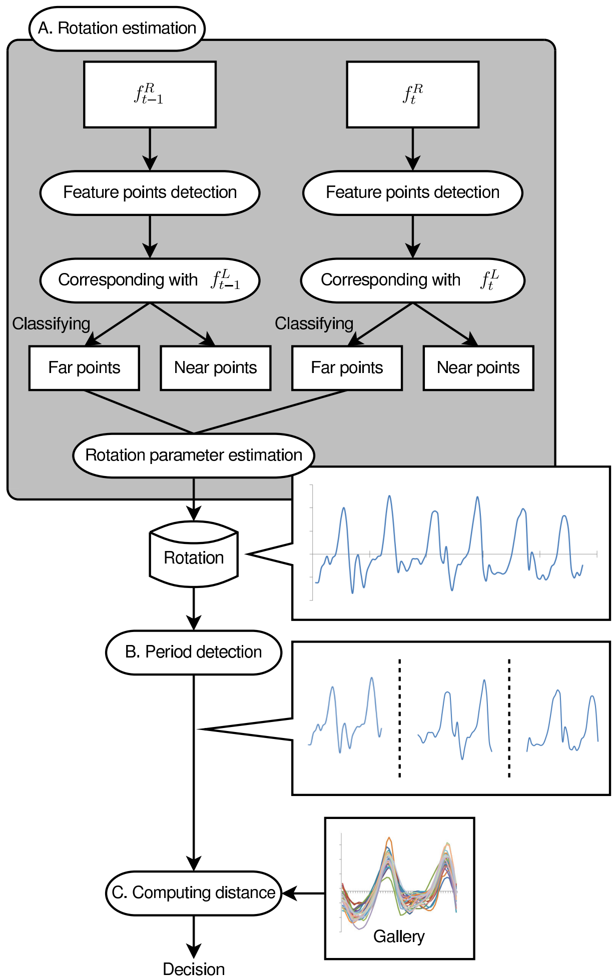


Fig. 3. The flow diagram of vision-based personal authentication. Our personal authentication consists of three steps: A.Rotation estimation, B.Period detection and C.Computing distance.

proposed a robust and fast method which separately estimates translation and rotation by classifying feature points as far or near. Although this method used a special compound omni-directional vision sensor, it can be implemented with a common stereo camera pair with classification of feature points. We adopt this principle as a way to acquire gait information by classifying feature points into far/near. This classification is done by estimating depths of feature points with stereo matching.

We describe each step of rotation estimation at a certain time t. First, feature points on continuous two frames of the right camera of the stereo pair (f_{t-1}^R ! $\$$ f_t^R) are detected. Here, the coordinates of feature points are represented on an unit hemisphere centering at the pinhole of the right camera. Next, disparities of feature points are computed by stereo matching with corresponding frames of the left camera (f_{t-1}^L ! $\$f_t^L$), and feature points are classified into far feature points or near feature points according to their disparities. Rotation is then estimated using only far feature points, with the following RANSAC[11] based algorithm.

Algorithm: Rotation Parameter Estimation

1) Iterate below processes N times ($i = 1...N$):
   a) Randomly select two points $\boldsymbol{P}$, $\boldsymbol{Q}$ on f_{t-1}^R and two points $\boldsymbol{P}'$, $\boldsymbol{Q}'$ on f_t^R, respectively
   b) Compute $\boldsymbol{R}_i$, which moves $\boldsymbol{P}$ to $\boldsymbol{P}'$ and $\boldsymbol{Q}$ to $\boldsymbol{Q}'$ as shown in Fig.4, using Eq. (1)
   c) Count the number of supporters of $\boldsymbol{R}_i$ as an evaluation value for $\boldsymbol{R}_i$
2) Choose $\boldsymbol{R}_i$ which have the most number of supporters as $\boldsymbol{R}_t$

$$\boldsymbol{R} = [\boldsymbol{P}' \ \boldsymbol{Q}' \ \boldsymbol{n}'][\boldsymbol{P} \ \boldsymbol{Q} \ \boldsymbol{n}]^{-1} \tag{1}$$

Here, $\boldsymbol{n} = \boldsymbol{P} \times \boldsymbol{Q}$ and $\boldsymbol{n}' = \boldsymbol{P}' \times \boldsymbol{Q}'$. A supporter is a pair of points ($\boldsymbol{X}$, $\boldsymbol{X}'$) which approximately satisfies $\boldsymbol{X}' = \boldsymbol{R}\boldsymbol{X}$. In other words, a rotation matrix with a small number of supporters has just a few number of correspondences of feature points between f_{t-1}^R and f_t^R, and a rotation matrix with a large number of supporters has a lot of correspondences. That is, the consensus of this algorithm is the rotation matrix with the most number of supporters in N tries.

Because there are only three degrees of freedom in the rotation matrix, we can represent $\boldsymbol{R}_t$ using roll-pitch-yaw angles $(\phi_t, \theta_t, \psi_t)$. Therefore, we define gait signal sequence in our method as $\boldsymbol{A} = \{\boldsymbol{a}_t\}$. Here, $\boldsymbol{a}_t = (\phi_t, \theta_t, \psi_t)$.

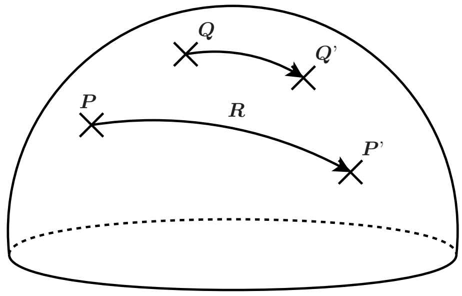


Fig. 4. Computing rotation between continuous two frames. $\boldsymbol{P}$ and $\boldsymbol{Q}$ are randomly selected feature points in f_{t-1}^R, $\boldsymbol{P}'$ and $\boldsymbol{Q}'$ are ones in f_t^R. $\boldsymbol{R}$ is a rotation matrix transforming $\boldsymbol{P}$ to $\boldsymbol{P}'$ and $\boldsymbol{Q}$ to $\boldsymbol{Q}'$.

B. Period detection

Next, the obtained gait signal sequence $\boldsymbol{A}$ is segmented into gait periods. Conventional methods [1][2][3][4] detect periods based on heuristic knowledge such as local peaks and valleys and/or auto-correlation of signals. As such, they sometimes fail to accurately detect periods, due to noise in the gait signal. In contrast, Makihara et al. [12] proposed a robust method for the period segmentation of a single quasi-periodic signal using Self Dynamic Time Warping (Self DTW). However, this method involves the problem of ambiguity, which is a combination of a phase evolution function and a periodic signal, producing temporally distorted signals compared to the original signals. Trung et al. [5] proposed a method for accurate period detection by applying [12] to a quasi-periodic signal of gait signal, reducing such ambiguity by linearization of Time Warping Function (TWF). Our method detects periods using [5]. A gait signal sequence $\boldsymbol{A}$ is segmented into gait periods $\boldsymbol{p}_j$, and a probe $\mathbb{P}$, which is a set of input periods, is constructed as $\mathbb{P} = \{\boldsymbol{p}_j\}$. For more technical details, please refer to [5] and [12].

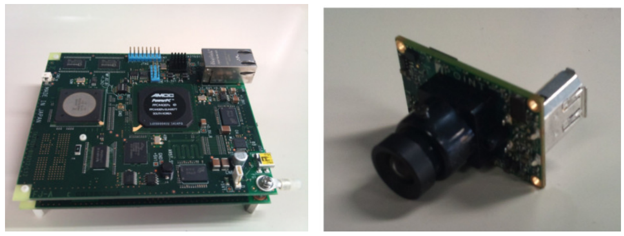
C. Computing distance and Authentication

The distance between a probe and a gallery is computed. We employ an elastic signal matching algorithm, Dynamic Time Warping (DTW) as with [13][5] to compute the distance. Because DTW is robust over temporal distortion, it is suitable for matching of gait signals which often suffer from temporal distortion due to variation in ground conditions, walking speed and so on.

A gallery $\mathbb{G} = \{\boldsymbol{g}_k\}$ is a collection of the owner's sample gait periods, constructed via period detection. Similarly, a probe $\mathbb{P} = \{\boldsymbol{p}_j\}$ is a collection of input

(a) appearance (b) inside

Fig. 5. The overview of the prototype system.



(a) NXV1-1394-PCB (b) FFMV-03MTC

Fig. 6. Components of the prototype system.

TABLE I
THE SPECIFICATIONS OF NXV1-1394-PCB.

CPU	PowerPC440EPX 666MHz
Flash Memory	128MB
RAM	256MB
OS	Linux 2.6.27.9
Image Processor	MB87Q0530
Image Memory	64MB
Dimensions	120mm $\times$ 100mm $\times$ 40mm
Weight	180g
Power	15W

periods. The distance between $\mathbb{G}$ and $\mathbb{P}$ is then computed.

$$Dist(\mathbb{G}, \mathbb{P}) = \min_{j,k} d(\boldsymbol{p}_j, \boldsymbol{g}_k). \tag{2}$$

Here, $d(\cdot, \cdot)$ denotes a normalized cumulative DTW score at the end of the optimal warping path [5][13].

In the last step, the computed distance between $\mathbb{G}$ and $\mathbb{P}$ is compared with the threshold. If the distance is less than the threshold, the wearer is authenticated as the owner.

III. PROTOTYPE SYSTEM

This section describes the prototype of the WS system. Fig. 5 shows the overview of the prototype system. This system is designed for school children, with the apparatus based on a schoolbag.

The computer in this system must be sufficiently compact to meet the requirement that it is easily wearable. In addition, it must have a certain level of computational capacity to deal with processor-heavy operations such as image processing. However, because there is generally a trade-off between portability and computing power, some hardware support is required to deal with image processing. As such, we adopt the NXV1-1394-PCB produced by FUJITSU Kyushu Network Technologies. The overview and the specifications of this computer are shown in Fig.6(a) and Table I. This computer contains an image processor MB87Q0530 in addition to its CPU. This image processor can perform high-speed image processing operation such as corner feature detection using Harris operator and template matching with normalized cross-correlation (NCC) algorithm. In addition, it is lightweight and uses little power, making it suitable for the WS system.

As a compact, lightweight stereo camera pair with low power consumption, we use the FFMV-03MTC, produced by Point Grey Research, Inc. An overview and the specifications of FFMV-03MTC are shown in Fig.6(b) and Table II. These cameras are installed on the top of the schoolbag (Fig.5(a)), to record in parallel stereo, and are connected to the computer inside the schoolbag. The cameras capture grayscale images at a sampling rate of 30 fps.

A portable battery supplies electricity to each component of the prototype system, with an estimated battery life of an hour and a half. This is a long enough period to cover common commuting times, which are often less than an hour. The total weight of the prototype system is approximately 600 g, not including the weight of the schoolbag. Thus, the system is light enough to be wearable.

We implemented rotation estimation, which is a part of our personal authentication method, in the prototype system. As described above, we applied hardware acceleration with the image processor in the computer for feature point detection and template matching. Feature points on f_{t-1}^R and f_t^R are detected using Harris operator in the image processor, as shown in Figs. 7(a) and 7(b). These points are classified into far and near points using template matching, as shown in Figs. 7(c) and 7(d). Far and near points are indicated by blue and red dots, respectively. Supporters between f_{t-1}^R and f_t^R are computed by processing on the CPU as shown in Figs. 7(e) and 7(f). They are indicated by pink dots.

Fig. 8 shows a sample of estimated rotation in real-

TABLE II
THE SPECIFICATIONS OF FFMV-03MTC.

Maximum Resolution	752 $\times$ 480
Pixel Size	6.0μm $\times$ 6.0μm
Dimensions	22.4mm $\times$ 44mm $\times$ 34mm
Weight	37g
Power	1W or less
Connector	FireWire

TABLE III
PROCESSING TIME FOR ROTATION ESTIMATION.

	Average processing time [ms]
Feature point detection	3.8
Far/near classification	3.7
Rotation estimation	22.5
Total	30.0

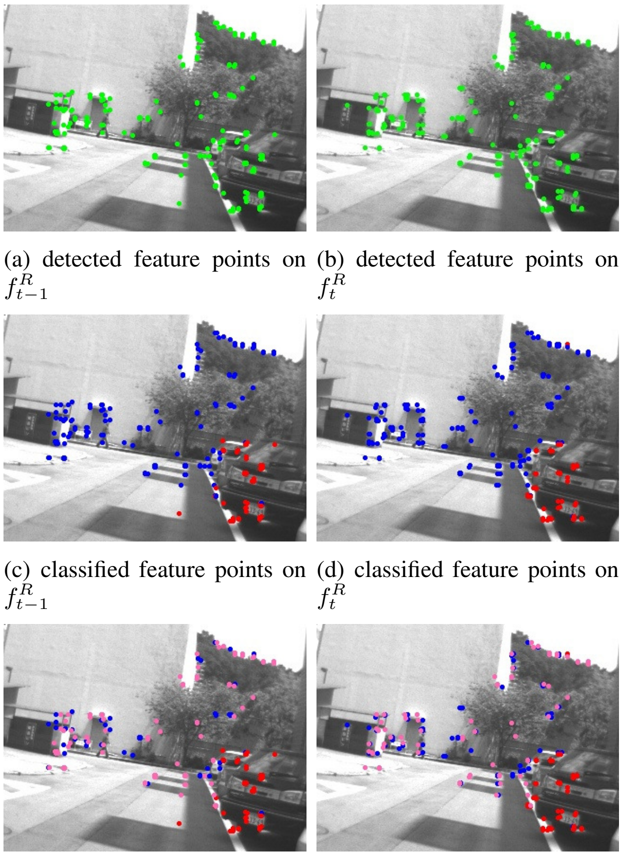
time, performed by the prototype system. From Fig. 8, it can be seen that quasi-periodic features appeared in the gait signals. Table III shows the average processing time for each step of the rotation estimation procedure. Using hardware acceleration by the image processor, feature point detection and far/near classification were executed in 3.8 ms and 3.7 ms, respectively. Rotation parameter estimation had a time cost 22.5 ms, and the total computational time was 30.0 ms. Because this total time was shorter than 33 ms, which is a sampling period at 30 fps, it can be seen that rotation is able to be estimated in real-time with the prototype system.

IV. EXPERIMENTS

An experiment was conducted to test the validity of the proposed method. To analyze the system's performance in more detail, we used images that were pre-captured offline. Rotation was estimated with the NXV1-1394-PCB, and other computations were performed offline.

We examined data from 39 adult subjects (32 male and 7 female). The subject's age varied between 21 and 50. Each subject walked at their normal walking speed in an outdoor environment. For each subject, data was recorded for four sequences of 30 seconds each. One sequence was used in a gallery; the others were used for the probe. To examine changes due to the duration of the probe, We measured probes that were 15 seconds or 8 seconds in length. These probes generated by cutting each 30 second sequence.

To evaluate the proposed method quantitatively, Receiver Operation Characteristic (ROC) curves in terms of



(a) detected feature points on f^R_{t-1} (b) detected feature points on f^R_t

(c) classified feature points on f^R_{t-1} (d) classified feature points on f^R_t

(e) supporters on f^R_{t-1} (f) supporters on f^R_t

Fig. 7. Samples of processed images. Feature points on f^R_{t-1} and f^R_t are detected using Harris operator in the image processor, as shown in (a) and (b). These feature points are classified into far and near points using template matching, as shown in (c) and (d). Far and near points are indicated by blue and red dots, respectively. Supporters between f^R_{t-1} and f^R_t are computed by processing as shown in (e) and (f). They are indicated by pink dots.

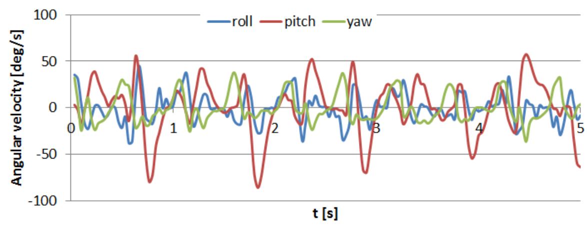
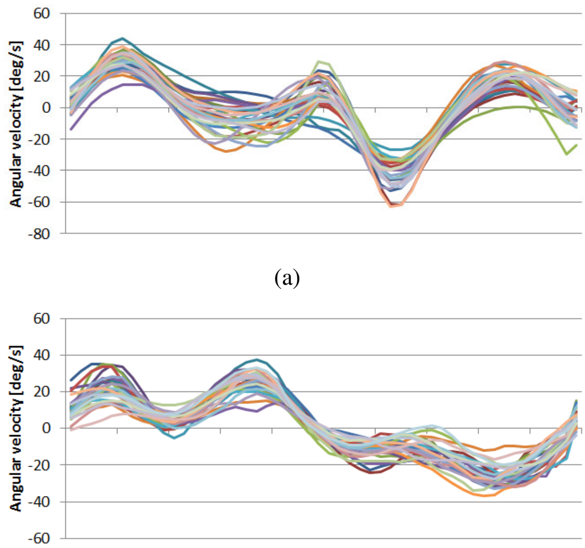


Fig. 8. A sample of estimated rotation: It can be seen that quasi-periodic features appeared in the gait signals.

False Acceptance Rate (FAR) and False Rejection Rate (FRR), and the Equal Error Rate (EER) were calculated.

Fig. 9 shows sample galleries, composed of gait periods by applying period detection to rotation, as in Fig. 8. Here, only signals of pitch angle are presented. Figs 9(a) and 9(b) show galleries generated by rotation



(a)

(b)

Fig. 9. Sample galleries. Only signals of pitch angle are presented. (a) and (b) show galleries generated by rotation from the gait of different people. Clear difference can be seen between (a) and (b).

TABLE IV
EQUAL ERROR RATE.

	sensor location	EER[%]	#S
Ailisto et al. [2]	waist	6.4	36
Mantyjarvi et al. [1]	waist	7-19	36
Gafurov et al. [3]	lower leg	5, 9	21
Trung et al. [5]	back-bag	6.0	32
Our method (15s)	back-bag	5.6	39
Our method (8s)	back-bag	6.8	39

from the gait of different people. Although there are a small number of outliers, the periods were well extracted overall. In addition, clear difference can be seen between Figs. 9(a) and 9(b).

Fig. 10 and Table IV show ROC curves and EER, respectively. Here, #S represents the number of subjects. In terms of EER, the results revealed that our method using wearable cameras exhibited comparable performance to existing methods using motion sensor signals. Furthermore, even when the length of the probe was shortened to 8 seconds, the results were reasonable. Table V shows the changes of average processing time for authentication process. Regardless of the length of probe, required time is shorter than one second. This is short enough because the time for input capture is 8 seconds or more.

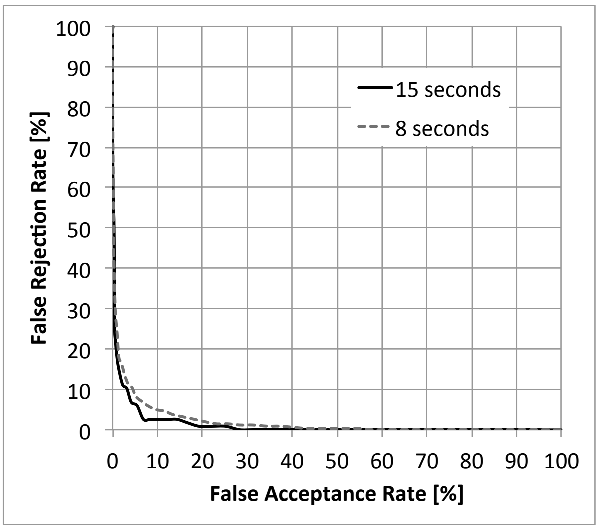


Fig. 10. Receiver Operating Characteristic curves.

TABLE V
PROCESSING TIME FOR AUTHENTICATION.

	Average processing time [s]
Our method (15s)	0.24
Our method (8s)	0.12

V. LIMITATION

Feature point detection: In our method, gait information was estimated depending on feature points in camera images. Therefore, if feature points are not stably detected (i.e. because of lighting conditions and motion blur etc.) our method would be unable to perform well.

VI. CONCLUSION AND FUTURE WORK

This paper proposed a gait-based person authentication method for a WS system. Our method, departing from previous methods using motion sensor signals, can authenticate users based on signals estimated between contiguous camera images. We tested our system experimentally, revealing EER values of 5.6% and 6.8%. This result indicates that our visual-based method could successfully identify individuals based on gait analysis. Our method exhibited comparable accuracy to previous methods.

In future work, we hope to further develop our method for practicable applications. In the WS system, all authentication processes need to be run on a compact computer in a stand-alone environment. However, this time both of period detection and distance computation processes are performed on a desktop PC. To provide our method as a function of the WS system, we have to seamlessly connect each step of the process of it. Moreover, in our future work we plan to implement the remaining functions mentioned in the first section: the detection of suspicious individuals nearby, and notifying another person of the situations. By these functions, the system becomes able to transmit images of their child's surroundings to parents via the network, and photos of any suspicious individuals, for example. The former function can be implemented using existing face detection/recognition techniques. To realize the latter function, we plan to utilize network equipments so the WS system can access a public communication network. After achieving these functions, the system will be evaluated by elementary school students.

We will conduct experiments on a real way to school by elementary school children. It will reveal whether our authentication method is available for children and whether our WS system is useful to increase safety.

ACKNOWLEDGEMENT

This work was supported by Grant-in-Aid for Scientific Research (S) 21220003.

REFERENCES

[1] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, march 2005, pp. ii/973 – ii/976 Vol. 2.

[2] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S.-M. Makela, *Identifying people from gait pattern with accelerometers*. SPIE, 2005, vol. 5779, no. 1, pp. 7–14.

[3] D. Gafurov, K. Helkala, and T. Sondrol, "Biometric gait authentication using accelerometer sensor," *Journal of Computers*, vol. 1, no. 7, 2006.

[4] D. Gafurov and E. Snekkenes, "Gait recognition using wearable motion recording sensors," *EURASIP Journal on Advances in Signal Processing*, no. August, pp. 1–17, 2009.

[5] N. T. Trung, Y. Makihara, R. S. Hajime Nagahara, Y. Mukaigawa, , and Y. Yagi, "Phase registration in a gallery improving gait authentication," *International Joint Conference on Biometrics*, Oct. 2011.

[6] M. Finotto and E. Menegatti, "Humanoid gait stabilization based on omnidirectional visual gyroscope," in *Proceedings of the 4th Workshop on Humanoid Soccer Robots*, December 2009, pp. 52–59.

[7] J. Gluckman and S. Nayar, "Ego-motion and omnidirectional cameras," in *Computer Vision, 1998. Sixth International Conference on*, jan 1998, pp. 999 –1005.

[8] Y. Yagi, W. Nishii, K. Yamazawa, and M. Yachida, "Rolling motion estimation for mobile robot by using omnidirectional image sensor hyperomnivision," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 1, aug 1996, pp. 946 –950 vol.1.

[9] R. C. Nelson and J. Aloimonos, "Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head)," *Biological Cybernetics*, vol. 58, pp. 261–273, 1988.

[10] N. T. Trung, Y. Kojima, R. S. Hajime Nagahara, Y. Mukaigawa, M. Yachida, and Y. Yagi, "Real-time estimation of fast egomotion with feature classification using compound omnidirectional vision sensor," *IEICE Transactions on Information and Systems*, vol. E93-D, no. 01, pp. 152—166, Jan. 2010.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[12] Y. Makihara, N. Trung, H. Nagahara, R. Sagawa, Y. Mukaigawa, and Y. Yagi, "Phase registration of a single quasi-periodic signal using self dynamic time warping," in *Computer Vision – ACCV 2010*, ser. Lecture Notes in Computer Science, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Springer Berlin / Heidelberg, 2011, vol. 6494, pp. 667–678, 10.1007/978-3-642-19318-7_52.

[13] M. Derawi, P. Bours, and K. Holien, "Improved cycle detection for accelerometer based gait authentication," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, oct. 2010, pp. 312 – 317.