

The Optimal Camera Arrangement by a Performance Model for Gait Recognition

Naoki Akae, Yasushi Makihara, and Yasushi Yagi
The Institute of Scientific and Industrial Research
Osaka University

8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, JAPAN
{akae,makihara,yagi}@am.sanken.osaka-u.ac.jp

Abstract—Recently, many gait recognition algorithms are proposed, and the optimal camera arrangement is necessary to maximize the performance. In this paper, we propose the optimal camera arrangement by using a performance model that considers observation conditions comprehensively. We select silhouette resolution, observation view, and its local and global changes as the observation conditions affecting the performance. Then, training sets composed of pairs of the observation conditions and the performance is obtained by gait recognition experiments under several camera arrangements. A performance model is constructed by applying Gaussian Processes Regression to the training set. The optimal arrangement is determined by estimating the performance for each camera arrangement with the performance model. The effectiveness of the proposed method is demonstrated by experiments of performance estimation with a training set including 17 subjects and the optimal camera arrangement.

I. INTRODUCTION

In the modern society, there is a growing need to identify individuals in many different situations, including surveillance and access control. For personal identification, many biometric-based authentication methods have been proposed using a wide variety of cues, such as fingerprint, iris, faces, and gait. Of these, gait identification has attracted considerable attention because it offers surveillance systems the ability to ascertain identity at a distance.

There are several observation conditions affecting gait recognition performance such as observation view and distance between the subjects and the camera. Since these observation conditions are dependent on the camera arrangement, the optimal camera arrangement is required for better gait recognition.

Currently, most of the modeling and analysis of gait recognition performance by observation conditions are focused on two points, namely, observation view and distance between the subjects and the camera (resolution).

As for the observation view, Yu et al. [1] modeled the relationship between the performance and the view angle on appearance-based gait recognition. Makihara et al. [2] analyzed reference views' effect on the performance for View Transformation Model (VTM). Wang et al. [3] proposed a multi-view fusion algorithm based on the Dempster-Shafer rule, and analyzed the performance in terms of the pair of different observation views. Sugiura et al. [4] used an omnidirectional camera to capture multi-view gait images form a

single sequence, and reported that the performance increased as the number of observation views increased.

As for the distance between the subjects and the camera (resolution), it was reported that the performance decreased as the resolution decreased[5][6].

There is, however, another observation conditions affecting the performance in addition to these two conditions. Currently, most appearance-based gait recognition algorithms use a period-based gait feature[7][8], and hence observation view changes in a gait period may degrade the performance because extracted gait features become unstable. Therefore, we should also consider the observation view change in a gait period as an observation condition.

Moreover, the previous analyses dealt with a part of the observation conditions separately, and there is no algorithms dealing with all the observation conditions comprehensively.

Therefore, we propose a method of the optimal camera arrangement based on a performance model dealing with all the observation conditions comprehensively. First, we obtain training sets composed of pairs of the observation conditions and the performance by gait recognition experiments under a limited number of conditions. Second, the performance model is constructed by regression analysis on the training sets. Finally, the optimal camera arrangement is determined by estimating the performance for a given camera arrangement based on the model.

The outline of this paper is as follows. First, assumption in this research is described in Section II. Next, the observation conditions affecting the performance are discussed in Section III, and the construction of the performance model which outputs the performance from the observation conditions is described in Section IV. Finally, the model and the optimal camera arrangement are evaluated in Section V. Section VI contains conclusions and discussion for further work.

II. ASSUMPTION

In this section, we introduce two assumptions.

Although there are several algorithms that improve the performance by using multiple cameras [3], it is undesirable to use them in terms of cost. Thus, we treat the optimal camera arrangement for a single camera.

Another assumption is related to walking courses. In case of access control system for example, we can assume that the walking courses of subjects are roughly pre-determined

and straight (e.g., the center course at the corridor). In this problem setting, once a camera arrangement is given, observation condition is also given and hence the performance can be estimated from a generic performance model constructed in a training environment in advance.

III. OBSERVATION CONDITIONS

In this section, we discuss the observation conditions affecting the performance and use the following five observation conditions.

- Observation view
- Observation view variation in sequence
- Observation view change in gait period
- Silhouette resolution of upper body
- Silhouette resolution of lower body

Table I shows the qualitative relationship between the performance and the four observation conditions except for observation view. In the following sections, these observation conditions are defined in world coordinate where subjects walk straight parallel to X axis, as illustrated in Fig. 1. The captured section is referred to a sequence, and its length is defined as L .

A. Observation view

Property of gait feature is dependent on the observation view. For example, while dynamic features such as step and arm swing are clearly seen in side view gait images, static features such as shape of body are clearly seen in front view gait images. Because the performance may change by such properties changes, we take observation view into account.

The observation view θ is defined by walking direction and azimuth view at the center point P_C in the sequence as illustrated in Fig. 1.

B. Observation view variation in sequence

As reported in [4], the larger the observation view variation in sequence is, the higher the performance is.

Observation view variation in sequence θ_{seq} is defined as the difference between θ_s and θ_e , that is, the observation views at the start and end point in the sequence.

Observation view variation in sequence is decided by the sequence length, the view angle of a camera, the distance between the camera and the subject, and observation view.

In case where an omni-directional camera is used as a sensor, we can capture front view, side view, and back view gait images. And observation view variation in sequence becomes the largest if the sequence length L is infinity. However, the sequence length L is often limited at an actual environment, and observation view variation in sequence is decided by the sequence length. In the case of a perspective camera as sensor, observation view variation is further reduced due to limitation of view angle of the perspective camera.

Moreover, even if the sequence length is the same, the smaller the distance between the camera and the subject is, the larger observation view variation in sequence is. In addition, observation view also affects the observation view variation, because it is large when a camera is arranged

TABLE I
EFFECT ON THE PERFORMANCE BY OBSERVATION CONDITIONS

Conditions	Performance	
	Large	Small
Observation view variation in sequence	High	Low
Observation view change in 1 gait period	Low	High
Silhouette resolution of upper body	High	Low
Silhouette resolution of lower body	High	Low

to capture near-side view images, while it is small when a camera is arranged for near-front view images.

C. Observation view change in gait period

If observation view in gait period is large, extracted period-based gait features get unstable and the performance decreases. Therefore, we take observation view change in gait period into consideration.

When multiple gait features are extracted from a sequence as illustrated in Fig. 2, observation view change in gait period θ_{sub} is calculated as average of observation view changes in each gait period.

$$\theta_{sub} = \frac{1}{N_{sub}} \sum_{i=1}^{N_{sub}} \theta_i, \quad (1)$$

where N_{sub} is the number of gait features extracted from the sequence and θ_i is observation view change in i -th gait period where i -th gait feature is extracted. Note that N_{sub} is dependent on the interval of center points P_i , and that we define the interval as 0.5 m empirically in this paper.

This factor, independent of sensor, changes by the distance between the camera and the subject, and by observation view in the same way as observation view variation in sequence.

D. Silhouette resolution

It is reported in [5][6] that low resolution silhouette results in low performance.

Silhouette resolution h_u and h_l are defined as upper and lower body heights in pixels in an input image respectively. Since silhouette resolution is different by body height of subjects, we use the resolution of average person whose upper body height is 85 cm and lower body height is 80 cm empirically.

The reason why we deal with upper and lower body resolution separately is (1) feature properties in upper and lower body are different and (2) their resolutions significantly differ, particularly in case where an omni-directional camera is used (resolution of central part of mirror is lower than that of peripheral part). In addition, resolution is in inverse proportion to the distance between the camera and the subject. As a result, silhouette resolution is dependent on the sensor and the distance.

E. Trade-off between observation conditions

There is generally no camera arrangements where all observation conditions have good effects on the performance,

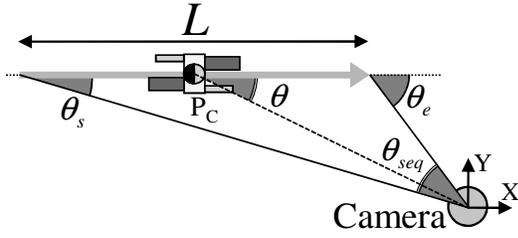


Fig. 1. Observation view and observation view variation in a sequence

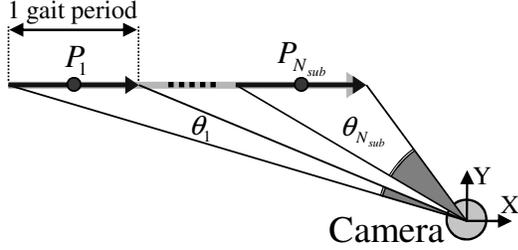


Fig. 2. Observation view change in gait period

namely, there is a trade-off among these observation conditions.

For example, if a camera is arranged at a far point from subjects, while silhouette resolution is low and observation view variation in sequence is small, observation view change in gait period is small and the extracted gait feature gets stable. On the other hand, if a camera is arranged at a near point from subjects, while silhouette resolution is high and observation view variation in sequence is large, observation view change in gait period is large and the extracted gait features gets unstable.

IV. PERFORMANCE MODEL

In this section, the performance model that outputs the performance from the observation conditions is constructed by Gaussian Processes Regression[9].

A. Definition of the performance

The performance is defined as Equal Error Rate (EER) on Receiver Operating Characteristics (ROC) curves[10]. The ROC curves shows relation between False Rejection Rate P_{FR} and False Acceptance Rate P_{FA} when the receiver changes the acceptance thresholds. EER is an error rate when P_{FR} becomes equal to P_{FA} , and smaller EER means higher performance. Note that EER is an example and that this performance model can deal with other evaluation standards such as rank-1 identification rate in the same way.

B. Uncertainty of the performance

In this section, the uncertainty of the performance (noise level) is modeled in order to introduce Gaussian Processes Regression. The uncertainty of the performance is generally dependent on the number of subjects and the number of attempts per subject, and the uncertainty decreases as these numbers increase. In this paper, the uncertainty of the performance is approximated by Gaussian noise that follows

a Gaussian distribution with zero mean and variance σ_i^2 , and σ_i^2 is estimated by the observed value of the performance, the number of subjects, and the number of attempts per subject. We approximately substitute the variance of false rejection rate for the variance of EER. The variance of false rejection rate $\hat{V}(P_{FR})$ is estimated as [11]

$$\hat{V}(P_{FR}) = \frac{1}{(N-1)} \left(\frac{\sum a_i^2}{M^2 N} - P_{FR}^2 \right), \quad (2)$$

where N is the number of subjects, M is the number of attempts per subject, and a_i is the number of false non-matches for the i -th subject.

C. Modeling by Gaussian Processes Regression

First, input vector is defined as $\mathbf{x} = (\theta, \theta_{seq}, \theta_{sub}, h_u, h_l)^T$, where each element is the observation condition defined in section III and the dimensions D of input vector \mathbf{x} is 5.

Next, n input vectors $\{\mathbf{x}^i\} (i = 1, \dots, n)$ are aggregated in a $D \times n$ matrix $X = (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^n)$, and n performance values $\{y^i\}$ are collected in the vector form $\mathbf{y} = (y^1, y^2, \dots, y^n)$, which are used as a training set $\mathcal{D} = (X, \mathbf{y})$. Therefore, the performance estimation problem is regarded as prediction of distribution of the performance $P(y_* | \mathbf{x}_*, \mathcal{D})$ by using Gaussian Processes Regression when the training set \mathcal{D} and a new input vector \mathbf{x}_* are given.

In this paper, we use Gaussian noise for the model as $y^i = f(\mathbf{x}^i) + \epsilon_i$, where y^i is the observed performance and is defined as the sum of function value f and additive noise ϵ_i . The variance of additive noise σ_i^2 is estimated by Eq. (2). By setting the variance for each input vector individually, data with small noise have a relatively greater effect on the estimated results and data with large noise have a smaller effect.

In order to deal with nonlinear regression, a regression function is defined as $f(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w}$, where ϕ is a function which maps a input vector \mathbf{x} into a high-dimensional feature space, and \mathbf{w} is a vector of regression parameters. Kernel function k that indicates the inner product of input vector \mathbf{x}^m and \mathbf{x}^n in the high-dimensional feature space is defined as

$$k(\mathbf{x}^m, \mathbf{x}^n) = \phi(\mathbf{x}^m)^T \phi(\mathbf{x}^n) \quad (3)$$

$$= v \exp \left[-\frac{1}{2} \sum_{d=1}^D \left(\frac{\mathbf{x}_d^m - \mathbf{x}_d^n}{r_d} \right)^2 \right], \quad (4)$$

where $\Theta = [v, r_1, \dots, r_D]$ is kernel function parameter set, which are estimated by minimizing the negative log likelihood $-\log p(\mathbf{y} | X, \Theta)$ with conjugate gradient method. The predictive distribution $P(f_* | \mathbf{x}_*, \mathcal{D})$ of the performance is defined as the following Gaussian distribution with mean \bar{f}_* and variance $V(f_*)$.

$$\bar{f}_* = \mathbf{k}_*^T (K + S)^{-1} \mathbf{y} \quad (5)$$

$$V(f_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (K + S)^{-1} \mathbf{k}_*, \quad (6)$$

where $K_{ij} = k(\mathbf{x}^i, \mathbf{x}^j)$, $\mathbf{k}_{*i} = k(\mathbf{x}_*, \mathbf{x}^i)$, $S_{ij} = \delta_{ij} \sigma_i^2$ and δ_{ij} is the Kronecker's delta.

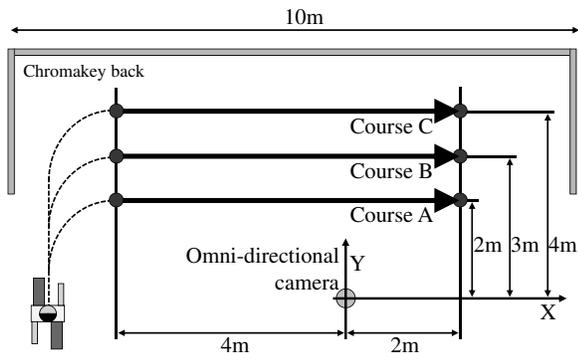


Fig. 3. Experimental environment

Then, the predictive distribution $P(y_*|\mathbf{x}_*, \mathcal{D})$ of performance with additive noise is defined as the following Gaussian distribution with mean \bar{y}_* and variance σ_*^2 .

$$\bar{y}_* = \bar{f}_* \quad (7)$$

$$\sigma_*^2 = V(f_*) + \sigma_{ob}^2, \quad (8)$$

where σ_{ob}^2 is the variance of observed noise estimated by \bar{y}_* and is approximated by estimated variance of false rejection rate $\hat{V}(P_{FR})$ as follows.

$$\sigma_{ob}^2 = \hat{V}(P_{FR}) = \frac{P_{FR}(1 - P_{FR})}{MN - 1} \quad (9)$$

Consequently, given a new input \mathbf{x}_* , performance y_* is estimated with the performance model.

V. EXPERIMENTS

In this section, the evaluation of the performance model and the optimal camera arrangement are described.

A. Datasets

All the subjects were asked to walk naturally along the three straight courses in an indoor environment as illustrated in Fig. 3. Each subject walked along each course 6 times over three days (twice per day). Thus, there are 18 gait sequences for each subject and 102 gait sequences for each course in our datasets. A test set is composed of 17 gallery sequences and 85 probe sequences for each course.

In this experiment, we used an omnidirectional camera (the camera was Point Grey Research GRS-50S5C-C), and images were captured by 1600×1200 pixel size and 20.3 ms shutter speed at 15 fps. Silhouette images were normalized by the silhouette height, and frequency domain feature[2] was used for gait recognition. Details of gait recognition method is described in [4].

B. Training sets

In order to construct an effective performance model, pairs of various observation conditions and EER are required. In this paper, various subsequences were extracted from these captured gait images by changing the subsequence length and the center point of subsequence.

First, gait features are extracted around center points lying at 0.5 m interval in range of $-3 \text{ [m]} \leq X \leq 1 \text{ [m]}$. Then, in order to change sequence length, the number for features used for recognition is set to 1, 4, and 7 respectively. Moreover, 1/2 and 1/4 times scaled images are also used for training. Next, all the subsequences provide pairs of observation conditions and EER through gait recognition experiments. Finally, performance model is trained by Gaussian Processes Regression using the training set.

C. Performance estimation

In this section, we evaluate the accuracy of estimated performance y_* . We prepare input vectors at 0.1 m interval on three courses in range of $-3 \text{ [m]} \leq X \leq 1.2 \text{ [m]}$. Moreover, the numbers of gait features for input vectors are set to 1, 2, 3, 4, 5, 6, and 7, and the scales of input images are set to 1, 3/4, 1/2, 3/8, and 1/4.

First, we evaluate the accuracy of Gaussian Processes Regression about two cases; (1) Sequence length is 1.5 m (single gait feature) and the scale of input image is 1.0, (2) Sequence length is 3.5 m and the scale of input image is 3/8. Figure 4 shows the results of regression for each course in case 1. From this graph, we can see that not only training data but also test data lie within estimated ranges. Figure 5 shows the results of regression for each course in case 2. Although all the input vectors in this case are excluded from training sets, we can see that the accuracy is sufficient.

Next, regression error normalized by standard deviation $(\bar{y}_* - y)/\sigma_*$ is shown in the histogram in Fig. 6. We can see that normalized regression errors at almost all the points are within the standard deviation range. Since average of normalized regression error is $0.25\sigma_*$, the accuracy of estimated y_* is sufficient.

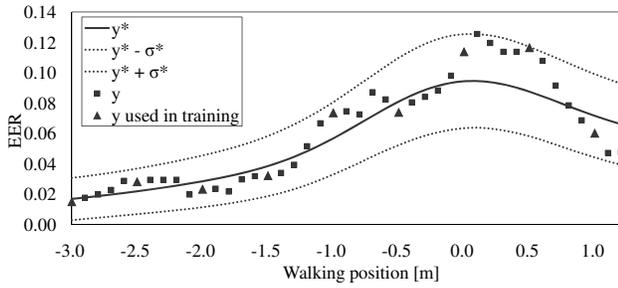
D. Optimal camera arrangement

In this section, we investigate the optimal camera arrangement by estimating the performance for each camera arrangement with the performance model.

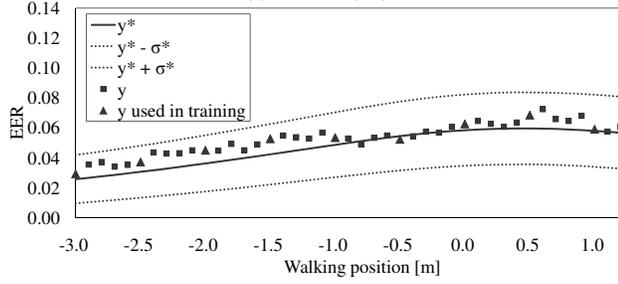
Omnidirectional camera: Figure 7 shows estimated EER and regression standard deviation $\sqrt{V(f_*)}$ in case of an omnidirectional camera.

As a result, regardless of the sequence length, EER is smaller when a camera is arranged so that front view gait images are captured. This means that observation view change in gait period have a large effect on the performance.

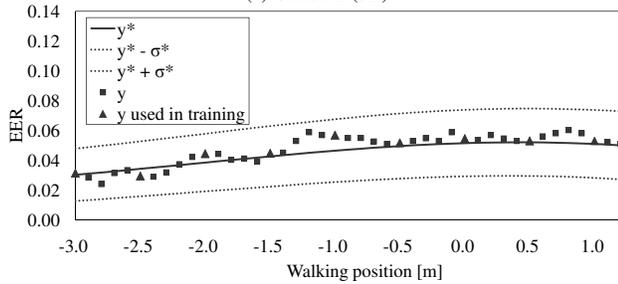
Moreover, we can see from the broken line that front view is the optimal observation view when depth is small, and that it gets closer to oblique view as depth increases. This result arises from a trade-off between observation view change in gait period and silhouette resolution as discussed in section III-E. In small-depth area, the observation view change in gait period in side or oblique view is quite large, and hence its effect overwhelms the effect of silhouette resolution. Therefore, the performance for near-front view is better than for side or oblique view. On the other hand, in large-depth area, observation view change in gait period sets small and their difference between the front view and the side



(a) Course A (2m)



(b) Course B (3m)



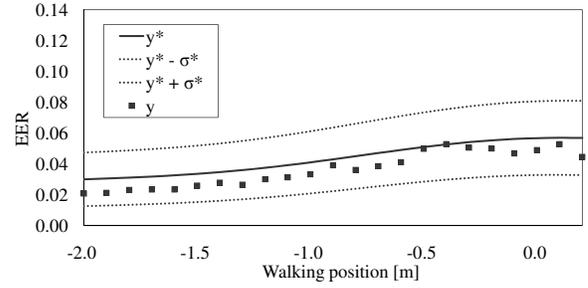
(c) Course C (4m)

Fig. 4. Regression result (Sequence length : 1.5 m, Scale : 1) The solid line means the estimated value \bar{y}_* by the performance model and the broken line means $\bar{y}_* \pm \sigma_*$.

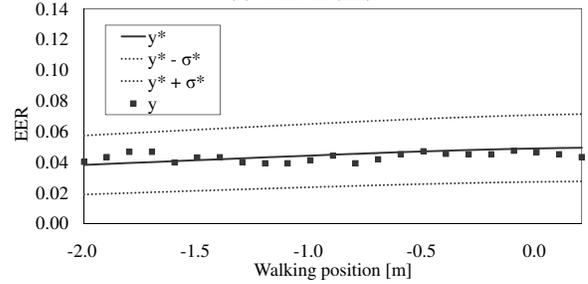
or oblique view becomes relatively small, while silhouette resolution in side or oblique view is larger than that of near-front view for the same depth. Therefore, the effect of the silhouette resolution becomes relatively larger than that of observation view change in gait period, and the performance for side or oblique view gets better than that of near-front view.

As the advantage of Gaussian Processes Regression, we can obtain the uncertainty of the estimated EER by the standard deviation, in other words, confidence domain where the performance model outputs reliable result to some extent. As a result, standard deviation is relatively small in the side or front oblique far area even though those areas are not included in training sets, and the reliability in the area is also relatively high. This is because scaled images are included in training sets.

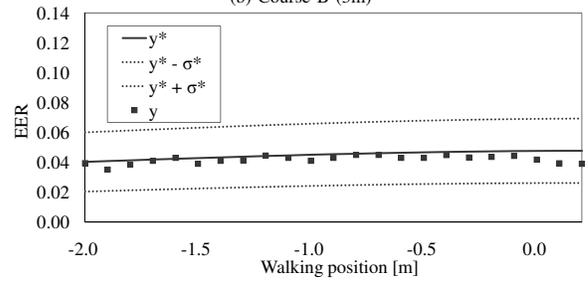
Perspective camera: Figure 8 shows estimated EER and regression standard deviation in case of a perspective camera whose view angle is 40 degrees. Since the view angle is limited, the sequence length gets shorter than the actual walking length particularly in case where the distance between the camera and the subject is small. We select appropriate focal



(a) Course A (2m)



(b) Course B (3m)



(c) Course C (4m)

Fig. 5. Regression result (Sequence length : 3.5 m, Scale : 3/8)

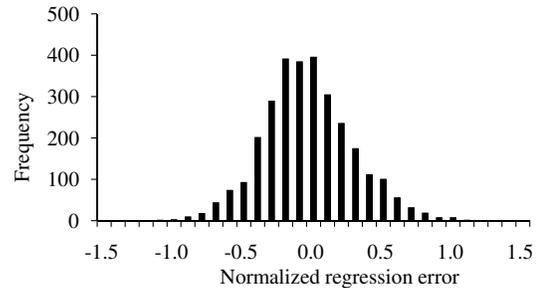


Fig. 6. Histogram of normalized regression error

length so that silhouette resolution is almost equal to the resolution in original scale of omni-directional images.

As a result, EER at the side view is larger than that of case of an omni-directional camera with 4.5m sequences (Fig. 7(e)). This is because the sequence length is shorter than the actual walking length due to the limitation of view angle.

Standard deviation is almost equal to the case of an omni-directional camera. This means that it is possible to estimate the performance with the almost same accuracy even if we use different types of cameras from the ones used in training.

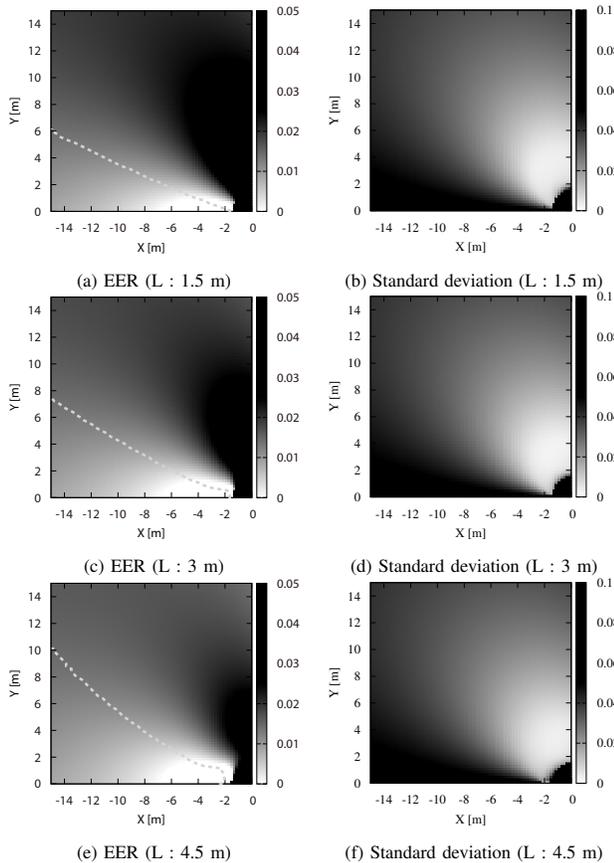


Fig. 7. Optimal arrangement (sensor : omni-directional camera) This figure shows estimated EER and regression standard deviation when an omni-directional camera is arranged at the point (0,0), subjects walk straight parallel to X axis, and (X, Y) is the center point of a sequence. If the estimated EER is negative, it is illustrated as 0. When the distance between the camera and the point (X, Y) is smaller than 1.5 m, EER and standard deviation are illustrated as 1 because whole body cannot be captured in the image. The broken line indicates the optimal points for each Y , namely, depth.

VI. CONCLUSION

This paper describes a method of optimal camera arrangement for gait recognition. First, we selected five observation conditions; observation view, observation view variation in sequence, observation view change in gait period, and silhouette resolutions for upper and lower body. Then, the performance model that outputs the performance from the five observation conditions was constructed by applying Gaussian Processes Regression to training sets. In the experiment, it is shown that almost all the estimated performances are within standard deviation. Finally, we analyzed the optimal camera arrangement by estimating the performance for each camera arrangement with the performance model.

Future work includes experiments in different environments from that of training set. In addition, extraction of stable gait feature from sequences with large observation view change in gait period is needed.

REFERENCES

[1] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Proc. of 7th Asian Conf. on*

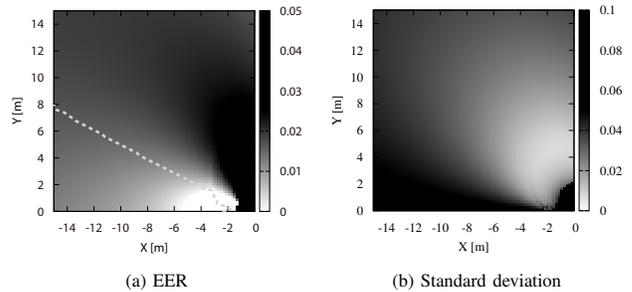


Fig. 8. Optimal arrangement (sensor : perspective camera, $L : 4.5$ m) This figure shows estimated EER and regression standard deviation when a perspective camera is arranged at the point (0,0), subjects walk straight 4.5m parallel to X axis, and (X, Y) is the center point of a sequence. If the estimated EER is negative, it is illustrated as 0. When a gait period of frames cannot be captured because of the small distance between the camera and the subject, EER and standard deviation are illustrated as 1. The broken line indicates the optimal points for each Y , namely, depth.

Computer Vision, volume 1, pages 807–816, Jan. 2006.

- [2] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Which reference view is effective for gait identification using a view transformation model? In *Proc. of the IEEE Computer Society Workshop on Biometrics 2006*, pages 1–8, New York, USA, Jun. 2006.
- [3] Y. Wang, S. Yu, Y. Wang, and T. Tan. Gait recognition based on fusion of multi-view gait sequences. In *Proc. of the IAPR Int. Conf. on Biometrics 2006*, pages 605–611, Jan. 2006.
- [4] Kazushige Sugiura, Yasushi Makihara, and Yasushi Yagi. Gait identification based on multi-view observations using omnidirectional camera. In *Proc. 8th Asian Conference on Computer Vision*, pages 452–461, Tokyo, Japan, Nov. 18-22 2007. LNCS 4843.
- [5] A. Mori, Y. Makihara, and Y. Yagi. Gait recognition using period-based phase synchronization for low frame-rate videos. In *Proc. of the 20th Int. Conf. on Pattern Recognition*, pages 2194–2197, Istanbul, Turkey, Aug. 2010.
- [6] S.D. Mowbray and M.S. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. In *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 566–573, 2003.
- [7] S. Sarkar, J.P. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.
- [8] J. Han and B. Bhanu. Individual recognition using gait energy image. *Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006.
- [9] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [10] P.J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. *Trans. of Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [11] A. J. Mansfield and J. L. Wayman. Best practices in testing and reporting performance of biometric devices, Aug. 2002.