Clothing-Invariant Gait Identification using Part-based Clothing Categorization and Adaptive Weight Control

Md. Altab Hossain, Yasushi Makihara, Junqiu Wang, Yasushi Yagi

Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047, JAPAN

Abstract

Variations in clothing alter an individual's appearance, making the problem of gait identification much more difficult. If the type of clothing differs between the gallery and a probe, certain parts of the silhouettes are likely to change and the ability to discriminate subjects decreases with respect to these parts. A part-based approach, therefore, has the potential of selecting the appropriate parts. This paper proposes a method for part-based gait identification in the light of substantial clothing variations. We divide the human body into 8 sections, including 4 overlapping ones, since the larger parts have a higher discrimination capability, while the smaller parts are more likely to be unaffected by clothing variations. Furthermore, as there are certain clothes that are common to different parts, we present a categorization for items of clothing that groups similar clothes. Next, we exploit the discrimination capability as a matching weight for each part and control the weights adaptively based on the distribution of distances between the probe and all the galleries. The results of the experiments using our large-scale gait dataset with clothing variations show that the proposed method achieves far better performance than other approaches.

Key words: Gait identification, Clothing-invariant, Part-based, Adaptive weight control, Biometrics.

1. Introduction

The demand for automated person identification systems is growing rapidly in many different areas, such as visual surveillance and access control. To effect person identification, many biometric-based identification methods have been proposed using a wide variety of cues, such as fingerprints, the face, palm, iris, voice, handwriting, and gait. Of these, gait has recently gained considerable attention as a promising biometric cue due to the possibility of identification at a distance without the subject's knowledge [1][2][3][4][5]. However, changes in appearance due to clothing variations, including coats, down jackets, baggy pants, and skirts, complicate the identification process.

Working from model-based approaches, Zhou et al. [6] proposed a method for human gait extraction based on a Bayesian framework using strong prior

knowledge. They applied it to person identification, allowing for variations in clothing such as skirts, by applying a pre-determined clothing model. The clothing type and its model are, however, not always available in advance in real applications.

Existing model-based gait silhouette cleaning methods [7][8] successfully remove carried objects or shadow regions as outliers. These silhouette cleaning techniques seem to be applicable to removing variations in clothing in gait identification. They are, however, prone to removing not only the carried objects and variations in clothing, but also individual differences which are crucial for person identification. In fact, it was reported in [7] that silhouette cleaning does not improve gait identification performance in the presence of carried objects.

With regard to appearance-based approaches, the LDA-based method [9] reduces to some extent the effects of intra-class variations, that is, clothing variations, on gait identification. It does not, however, work well when clothing variations exceed individual variations. Lee et al. [10] proposed a Shape Variation Based (SVB) frieze pattern as a gait feature for robust recognition. Because the feature is obtained by projecting a 2-D silhouette onto a 1-D horizontal or vertical axis, meaningful 2-D position information is degraded in the feature. In [11], self-similarity plots were used in gait recognition allowing for only a very limited change in clothing.

We have come to the conclusion that different clothing combinations generally affect different parts of the body. For example, consider a regular pair of pants and a full shirt as being the standard clothes of a subject. If the subject wears a hat and down jacket in addition to the standard clothes, it only affects the head and torso, respectively, and the other body parts remain unchanged and useful for recognition. In the event that a skirt and full shirt are worn instead, only the lower part of the body changes, allowing the other parts to be used for successful recognition. Therefore, a part-based approach has the potential of selecting the appropriate parts to be used for recognition.

One of the first methods that attempted to divide the human body into components for gait identification is described in [12]. The components are treated separately and applied in both person identification and gender classification. Experiments were, however, conducted with a relatively small data set. Boulgouris [13] proposed a component-based gait recognition that considers the unequal discrimination ability of each part. The method assigns fixed weights for the parts identified in the training phase. In [14][15], 7 gait components are defined, the contributions of which have been studied both individually and in certain combinations for both human gait recognition and gender recognition.

In the light of the above, we first constructed our own large-scale gait dataset with variations in clothing, and consisting of 68 subjects with at most 32 combinations of different types of clothing. We then proposed a part-based adaptive weight control for clothing-invariant gait identification. We divide the human body into 8 parts, including 4 overlapping parts to overcome the difficulties caused by different types of clothing. Body parts of varying lengths are considered, with the larger parts having a higher discrimination capability, while the smaller parts are more likely to be unaffected by variations in clothing. As there

are some similar clothes for each part, we have established a clothing categorization to group similar clothes. In the training phase, Probability Distribution Functions (PDFs) between gait features are stored separately for each part and for both the same and different clothes, and the corresponding discrimination capabilities are also measured in advance. In the test phase, given a probe, posteriors with the same and different clothes, are calculated based on the distances between the probe and the galleries and on the trained PDFs. Then a matching weight is calculated adaptively as the probability-weighted discrimination capability to reflect the importance of each part.

The outline of this paper is as follows. The construction of the large-scale dataset incorporating variations in clothing is introduced in Section 2. In Section 3, gait feature extraction is described briefly. Thereafter, gait identification based on adaptive weight control is addressed in Section 4. Experimental results using the dataset are presented in Section 5. Finally, our conclusions and future works are given in Section 6.

2. Gait dataset with variations in clothing

Early approaches used relatively small databases. Naturally, the success and evolution of a new application relies largely on the dataset used for evaluation. As gait is a behavioural biometric, there is plenty of scope for intra-subject variation, including clothing.

Previous gait datasets were limited in terms of subjects or clothing variations. The small Soton database [16] includes some clothing variations such as trainers, rain coats, and trench coats; however, there are only 12 subjects. On the other hand, the large Soton database [16], the HumanID dataset [17], and the CASIA dataset [18] have more than 100 subjects, but the variations in clothing are limited to casual wear and a long coat. Therefore, we constructed our own large-scale gait dataset with varying clothing. It includes 68 subjects with at most 32 combinations of types of clothing. Table 1 lists the clothing types, while Table 2 gives the combinations of clothing used in constructing the dataset. Figure 1 shows sample images of all the combinations of clothing types. All the gait sequences were captured twice on the same day in an indoor environment. Thus the total number of sequences in the dataset is 2746. The age distribution of the subjects in the dataset is shown in Fig. 2. The large number of subjects and clothing-variations in the new dataset provides us with an estimate of intra-subject variations together with inter-subject variations for a better assessment of the potential of gait identification.

3. Feature extraction

3.1. Part-based features

We divide the human body into 8 sections based on known anatomical properties [19]. For a body height H, the human body is segmented according to the vertical position of the neck ($0.87H$), waist ($0.535H$), pelvis ($0.48H$), and

Table 1: List of clothes used in the dataset (Abbreviation: name)

RP: Regular Pants	HS: Half Shirt	CW: Casual Wear
BP: Baggy Pants	FS: Full Shirt	RC: Rain Coat
SP: Short Pants	LC: Long Coat	Ht: Hat
Sk: Skirt	Pk: Parker	Cs: Casquette Cap
CP: Casual Pants	DJ: Down Jacket	Mf: Muffler

Table 2: Different clothing combinations

#	s_1	s_2	s_3	#	s_1	s_2	#	s_1	s_2
2	RP	HS	-	A	RP	Pk	T	Sk	FS
3	RP	HS	Ht	B	RP	DJ	U	Sk	Pk
4	RP	HS	Cs	I	BP	HS	V	Sk	DJ
9	RP	FS	-	K	BP	FS	D	CP	HS
X	RP	FS	Ht	J	BP	LC	F	CP	FS
Y	RP	FS	Cs	L	BP	Pk	E	CP	LC
5	RP	LC	-	M	BP	DJ	G	CP	Pk
6	RP	LC	Mf	N	SP	HS	H	CP	DJ
7	RP	LC	Ht	Z	SP	FS	0	CP	CW
8	RP	LC	Cs	P	SP	Pk	R	RC	RC
C	RP	DJ	Mf	S	Sk	HS			

#: Clothing combination type, s_i: ith clothes slot.

knee ($0.285H$), thereby defining the 8 parts shown in Fig. 3. The first part covers from the top of the head to the neck; the second part from the neck to the waist; the third part from the waist to the knee; the fourth part from the knee to the floor; the fifth part from the neck to the pelvis; the sixth part from the pelvis to the floor; the seventh part from the neck to the knee; and the final part from the top of the head to the pelvis. Note that the latter 4 parts overlap to overcome the difficulties created by clothing variations. The discrimination capability of the larger parts is high, but the probability of these being free of clothing differences is low, and vice versa for the smaller parts. We have thus included overlapping parts of varying lengths to minimize this tradeoff.

3.2. Frequency-domain features

In this section, frequency-domain feature extraction is addressed briefly (see [20] for a more detailed account). Given an image sequence, extracted silhouettes are size-normalized and registered to obtain a spatio-temporal Gait Silhouette Volume (GSV). Then, the gait period N_{gait} is detected by autocorrelation of the GSV, and the amplitude spectra at each pixel are calculated by Discrete Fourier Transformation (DFT) based on the gait period N_{gait} as shown in Fig. 4. In this paper, the image size of the GSV is 128×88, and frequencies between 0- and 2-times are used. Thus the gait feature vector composed of all the amplitude spectra is a $128 \times 88 \times 3 = 33,792$ dimensional vector. The low frequency

Figure 1: Sample clothes images

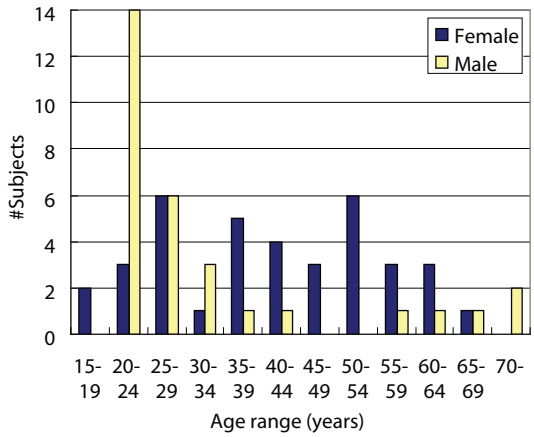


Figure 2: Age distribution of subjects in the dataset

elements are chosen because the high frequency elements contain noise rather than meaningful information and they do not improve gait identification performance. The meaning of each frequency can be explained as follows; the 0-times frequency represents the average silhouette, 1-time represents the asymmetry of the left and right motion, while the 2-times components represent the symmetry thereof. These features include both static (e.g. body shape, walking posture) and dynamic (e.g. stride, arm swing) components.

All parts are trained separately in the PCA subspace using the frequency-domain features from the corresponding regions, and dimension-reduced features are used as part-based gait features.

4. Part-based adaptive weight control

4.1. Overview

In a part-based method, the main issue is how to combine the individual distances into a single distance that quantifies the overall matching measure between a probe and the gallery. We adopt the weighted sum of distances corresponding to each part as the overall matching measure and adaptively control the weights for better identification, by increasing them for the same part of clothing and vice versa. In the following paragraphs, we give an overview of this adaptive weight control process.

Let $\boldsymbol{x}^p$ and $\boldsymbol{x}^g$ be sequences of a probe and the gallery, respectively. The matching measure between them is

$$D(\boldsymbol{x}^p, \boldsymbol{x}^g; \boldsymbol{X}^G) = \frac{\sum_{i=1}^{N_{parts}} w_i(\boldsymbol{x}^p; \boldsymbol{X}^G) d_i^z(\boldsymbol{x}^p, \boldsymbol{x}^g; \boldsymbol{X}^G)}{\sum_{i=1}^{N_{parts}} w_i(\boldsymbol{x}^p; \boldsymbol{X}^G)}, \tag{1}$$

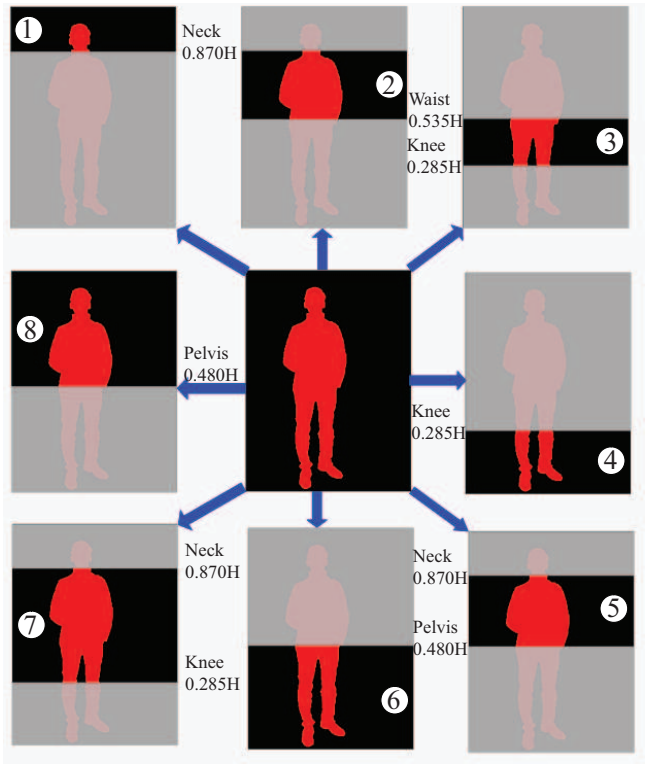


Figure 3: Definition of body parts

where N_{parts} is the number of body parts, $d_i^z(\boldsymbol{x}^p, \boldsymbol{x}^g; \boldsymbol{X}^G)$ is the z-normalized distance between $\boldsymbol{x}^p$ and $\boldsymbol{x}^g$ for the ith body part and $\boldsymbol{X}^G$ is the set of sequences of all the galleries used for z-normalization and for the subsequent probability calculation, and $w_i(\boldsymbol{x}^p; \boldsymbol{X}^G)$ is the weight of the ith body part. The weight $w_i(\boldsymbol{x}^p; \boldsymbol{X}^G)$ is the expected Fisher ratio and is expressed as the sum of the probability weighted Fisher ratios for the same and different clothes according to the following equation:

$$w_i(\boldsymbol{x}^p; \boldsymbol{X}^G) = P_i^{Sc}(\boldsymbol{x}^p; \boldsymbol{X}^G) F_i^{Sc} + P_i^{Dc}(\boldsymbol{x}^p; \boldsymbol{X}^G) F_i^{Dc}, \tag{2}$$

where F_i^{Sc} is the discrimination capability and $P_i^{Sc}(\boldsymbol{x}^p; \boldsymbol{X}^G)$ the posterior in the case that the clothes in the probe are included in the gallery clothing categories (let this case be Sc) for a given $\boldsymbol{x}^p$, and $\boldsymbol{X}^G$; F_i^{Dc} and $P_i^{Dc}(\boldsymbol{x}^p; \boldsymbol{X}^G)$ are, respectively, the same in the other case (let this case be Dc).

The discrimination capabilities and PDFs for the same and different clothes are learned during the training phase. Then, for a given probe $\boldsymbol{x}^p$, the posterior is calculated based on the learned PDFs and the gallery set $\boldsymbol{X}^G$. If the clothing type for the ith body part of the probe is similar to that in the gallery, the same clothes posterior $P_i^{Sc}(\boldsymbol{x}^p; \boldsymbol{X}^G)$ will be higher than the different clothes posterior

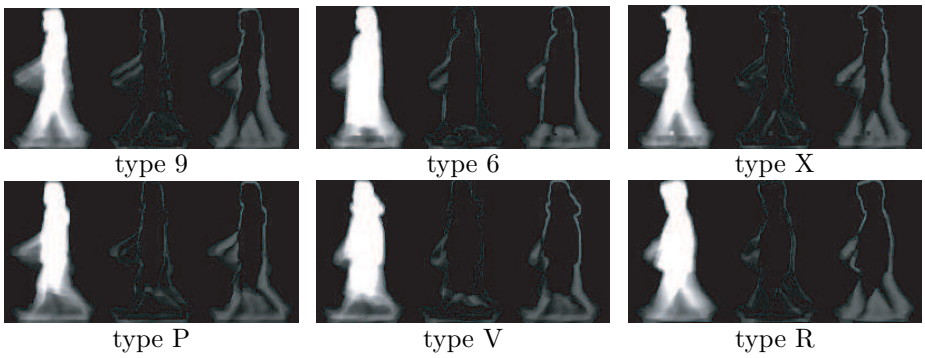


Figure 4: Frequency-domain gait features

$P_i^{Dc}(\boldsymbol{x}^p; \boldsymbol{X}^G)$ and vice versa. Thus the final weight obtained by combining the posterior and discrimination capabilities (Eq. (2)) is further increased for the part with the same clothes and vice versa, because, intuitively, F_i^{Sc} is greater than F_i^{Dc}. Note that the weights change adaptively for the probe and gallery sets. In the following subsections, we formulate each component.

4.2. Part-based matching measure

In this section, the part-based matching measure is defined. We define 4 alternative options for the matching measure. The first option is the mean value of the minimum distances of the subsequences for each probe and gallery. The second is the median value of the minimum distances of the subsequences for each probe and gallery. The third is the minimum value of the minimum distances of the subsequences for each probe and gallery. As a final option, we propose a new technique using the average of the 2 minimum distances of the subsequences for each probe and gallery.

Let $\{x_j^p\}(j = 1, 2, \ldots)$ and $\{x_k^g\}(k = 1, 2, \ldots)$ be subsequences with N_{gait} frames for the probe $\boldsymbol{x}^p$ and gallery $\boldsymbol{x}^g$, respectively. Let $\boldsymbol{a}_i(x)$ be the dimension-reduced feature vector for the ith body part for a subsequence x. The matching measure for the subsequences (let this be $d_i^{sub}(x_j^p, x_k^g)$) is simply chosen as the Euclidean distance between feature vectors $\boldsymbol{a}_i(x_j^p)$ and $\boldsymbol{a}_i(x_k^g)$. Then, the matching measure between the whole sequence for the ith body part is defined as the mean value of the minimum distances of the subsequences for each probe and gallery:

$$d_i(\boldsymbol{x}^p, \boldsymbol{x}^g) = \frac{1}{\sum_j 1} \sum_j [\min_k \{d_i^{sub}(x_j^p, x_k^g)\}] \tag{3}$$

Finally, the z-normalized distance is defined as

$$d_i^z(\boldsymbol{x}^p, \boldsymbol{x}^g; \boldsymbol{X}^G) = \frac{d_i(\boldsymbol{x}^p, \boldsymbol{x}^g) - \mu_i(\boldsymbol{x}^p; \boldsymbol{X}^G)}{\sigma_i(\boldsymbol{x}^p; \boldsymbol{X}^G)}, \tag{4}$$

where $\mu_i(\boldsymbol{x}^p; \boldsymbol{X}^G)$ and $\sigma_i(\boldsymbol{x}^p; \boldsymbol{X}^G)$ are, respectively, the mean and standard deviation of the distances between a probe $\boldsymbol{x}^p$ and a set of all the galleries $\boldsymbol{X}^G$ for the ith body part.

The second option for matching measure is the median value of the minimum distances of the subsequences for each probe and gallery:

$$d_i^{Med}(\boldsymbol{x}^p, \boldsymbol{x}^g) = Median_j[\min_k\{d_i^{sub}(x_j^p, x_k^g)\}] \tag{5}$$

For many probability distributions, the median gives a measure that is more robust in the presence of outlier values than the mean.

However, the distances calculated here have different properties. As the gallery sequences are similar, there is the possibility that the distances between the gallery and probe are biased towards the larger distances in real situations. The actual distances between galleries and probes can be obtained if there is no noise, segmentation errors or alignment problems. In practical applications, however, there is no way to overcome all the errors. The outliers tend to have larger values because the silhouettes in the galleries and probes contain errors. Therefore, the true values of the distances can be found in the smaller values and the median is not the best choice [21] in such a case. Choosing the minimum is a reasonable option which sometimes gives the best result. However, there is no guarantee that the minimum will always work in complex cases [1]. As a tradeoff, we propose a further option that uses the average of the 2 minimum values of the subsequences for each probe and gallery. The average of the 2 minimums can be thought of as a smoothing process. The minimum distances of the subsequences for the probe and gallery are calculated as:

$$d_{i,j}^{Min}(\boldsymbol{x}^p, \boldsymbol{x}^g) = \min_k\{d_i^{sub}(x_j^p, x_k^g)\} \tag{6}$$

Then, by sorting $d_{i,j}^{Min}(\boldsymbol{x}^p, \boldsymbol{x}^g)$ into ascending order, we get the first 2 values as the 2 minimum values, say $d_i^{min1}(\boldsymbol{x}^p, \boldsymbol{x}^g)$ and $d_i^{min2}(\boldsymbol{x}^p, \boldsymbol{x}^g)$. Finally, we compute the average of the 2 minimum values as the matching distance:

$$d_i^{Min2gp}(\boldsymbol{x}^p, \boldsymbol{x}^g) = \frac{1}{2}[d_i^{min1}(\boldsymbol{x}^p, \boldsymbol{x}^g) + d_i^{min2}(\boldsymbol{x}^p, \boldsymbol{x}^g)] \tag{7}$$

We have compared the performance of all 4 options for the matching measure (Eq. (3), Eq. (5), Eq. (6), and Eq. (7)) using different gallery and probe sets. The results show that the measure using Eq. (7) gives the best results, while Eq. (3) is the worst. For the details of the comparison, see Section 5.2. We have therefore chosen the average of the 2 minimum values, Eq. (7), as our matching measure.

4.3. Clothing categorization for each part

There are some similar clothing types for each part. For example, the standard clothes (RP + FS) and clothing type T (Sk + FS) are the same for the 2nd body part (the torso), but different for the 4th body part (the leg), and vice

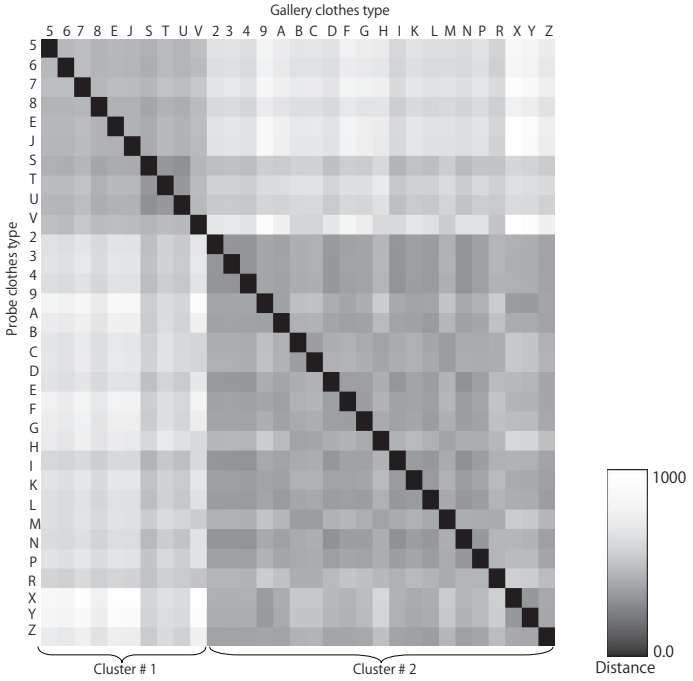


Figure 5: Distance matrix (sorted by cluster label) showing the clothing clustering for the 3rd body part

versa in the case of the standard clothes (RP + FS) and type B (RP + DW). Thus to define "the same clothes" and "different clothes" for each specific body part in Eq. (2), clothing should be appropriately categorized for each part.

Clothing types are treated separately as initial clusters. The distance between clusters is defined by averaging the distances between all the subjects with the relevant clothing types. The clothing category is obtained from a hierarchical agglomerative algorithm based on Ward's method [22]. Then, as clustering progresses, similar clothing types are merged. Figure 5 shows the clustering result for the 3rd body part.

The optimal cluster number is chosen so as to maximize the Fisher ratio between the PDFs of distance for the same and different clothes. Fisher's ratio is calculated for each cluster number (for each hierarchy of clustering) as shown in Fig. 6. Fisher's discriminant ratio is defined as

$$F_i^N = \frac{(\sigma_{B,i}^N)^2}{(\sigma_{B,i}^N)^2 + (\sigma_{W,i}^N)^2} \tag{8}$$

$$(\sigma_{B,i}^N)^2 = w_{Sc,i}^N w_{Dc,i}^N (\mu_{Sc,i}^N - \mu_{Dc,i}^N)^2 \tag{9}$$

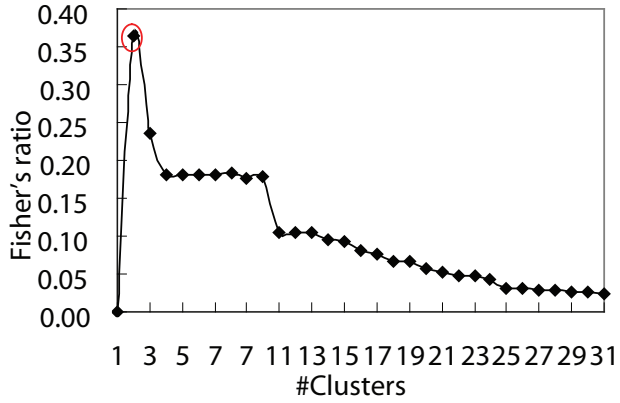


Figure 6: The Fisher ratios for cluster selection (3rd body part)

$$(\sigma_{W,i}^{N})^2 = w_{Sc,i}^{N}(\sigma_{Sc,i}^{N})^2 + w_{Dc,i}^{N}(\sigma_{Dc,i}^{N})^2 \tag{10}$$

$$w_{Sc,i}^{N} = \frac{n_{Sc,i}^{N}}{n_{Sc,i}^{N} + n_{Dc,i}^{N}} \tag{11}$$

$$w_{Dc,i}^{N} = 1 - w_{Sc,i}^{N}, \tag{12}$$

where N is the cluster number, $n_{Sc,i}^{N}$ is the number of pairs of the same clothes in the probe and gallery, $n_{Dc,i}^{N}$ is the number of pairs of different clothes in the probe and gallery, $(\sigma_{B,i}^{Sc})^2$ is the inter-class variance, and $(\sigma_{W,i}^{Sc})^2$ is the intra-class variance.

The number of clusters is selected based on the following criteria

$$^{\star}N_i = arg\max_{N} F_i^N \tag{13}$$

where $^{\star}N_i$ is the optimal number of clusters for the ith body part.

In Fig. 6, the number of selected clusters is 2 for the 3rd body part. The clothes are divided into 2 categories: (i) 5, 6, 7, 8, J, E (long coat) and S, T, U, V (skirt), and (ii) the remaining types.

4.4. Discrimination capability

Here we define the discrimination capability between the same and different subjects in Eq. (2) for the same and different clothes. In deriving these terms, suppose that a training set composed of pairs of probes and galleries is given and that it is divided into 4 sets for the ith body part:

- $U_{Ss,i}^{Sc}$ (the same subject with the same clothes),

- $U^{Sc}_{Ds,i}$ (different subjects with the same clothes),
- $U^{Dc}_{Ss,i}$ (the same subject with different clothes),
- $U^{Dc}_{Ds,i}$ (different subjects with different clothes).

We can define $U^{Sc}_{Ss,i}$ as

$$U^{Sc}_{Ss,i}=\{(\boldsymbol{x}^p,\boldsymbol{x}^g)|\boldsymbol{x}^p\in\boldsymbol{X}^P,\boldsymbol{x}^g\in\boldsymbol{X}^G,s(\boldsymbol{x}^p)=s(\boldsymbol{x}^g),c(\boldsymbol{x}^p)\in\boldsymbol{C}_i(\boldsymbol{x}^g)\}, \tag{14}$$

where $s(\boldsymbol{x})$ and $c(\boldsymbol{x})$ are the subject ID and clothing type, respectively, for a sequence $\boldsymbol{x}$, $\boldsymbol{C}_i(\boldsymbol{x})$ is the set of all the clothing types in the clothing category of the ith body part to which clothing type $c(\boldsymbol{x})$ belongs and $\boldsymbol{X}^P$ and $\boldsymbol{X}^G$ are sets of sequences of all the probes and galleries, respectively. The other sets $U^{Sc}_{Ds,i}$, $U^{Dc}_{Ss,i}$, and $U^{Dc}_{Ds,i}$ are defined in the same way.

$$U^{Sc}_{Ds,i}=\{(\boldsymbol{x}^p,\boldsymbol{x}^g)|\boldsymbol{x}^p\in\boldsymbol{X}^P,\boldsymbol{x}^g\in\boldsymbol{X}^G,s(\boldsymbol{x}^p)\neq s(\boldsymbol{x}^g),c(\boldsymbol{x}^p)\in\boldsymbol{C}_i(\boldsymbol{x}^g)\} \tag{15}$$

$$U^{Dc}_{Ss,i}=\{(\boldsymbol{x}^p,\boldsymbol{x}^g)|\boldsymbol{x}^p\in\boldsymbol{X}^P,\boldsymbol{x}^g\in\boldsymbol{X}^G,s(\boldsymbol{x}^p)=s(\boldsymbol{x}^g),c(\boldsymbol{x}^p)\notin\boldsymbol{C}_i(\boldsymbol{x}^g)\} \tag{16}$$

$$U^{Dc}_{Ds,i}=\{(\boldsymbol{x}^p,\boldsymbol{x}^g)|\boldsymbol{x}^p\in\boldsymbol{X}^P,\boldsymbol{x}^g\in\boldsymbol{X}^G,s(\boldsymbol{x}^p)\neq s(\boldsymbol{x}^g),c(\boldsymbol{x}^p)\notin\boldsymbol{C}_i(\boldsymbol{x}^g)\} \tag{17}$$

Based on the above definition, the following statistics are calculated,

$$N^{Sc}_{Ss,i}=\sum_{(\boldsymbol{x}^p,\boldsymbol{x}^g)\in U^{Sc}_{Ss,i}}1 \tag{18}$$

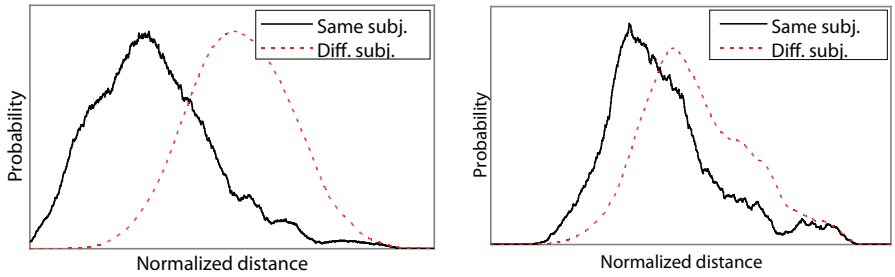
$$\mu^{Sc}_{Ss,i}=\frac{1}{N^{Sc}_{Ss,i}}\sum_{(\boldsymbol{x}^p,\boldsymbol{x}^g)\in U^{Sc}_{Ss,i}}d_{z,i}(\boldsymbol{x}^p,\boldsymbol{x}^g;\boldsymbol{X}^G) \tag{19}$$

$$(\sigma^{Sc}_{Ss,i})^2=\frac{1}{N^{Sc}_{Ss,i}}\sum_{(\boldsymbol{x}^p,\boldsymbol{x}^g)\in U^{Sc}_{Ss,i}}(d_{z,i}(\boldsymbol{x}^p,\boldsymbol{x}^g;\boldsymbol{X}^G)-\mu^{Sc}_{Ss,i})^2, \tag{20}$$

where $N^{Sc}_{Ss,i}$, $\mu^{Sc}_{Ss,i}$, and $\sigma^{Sc}_{Ss,i}$ are the number of pairs of a probe and gallery, and the average and variance of z-normalized distances between a probe and a gallery in set $U^{Sc}_{Ss,i}$, respectively. The statistics for the other sets are calculated in the same way.

Next, we formulate the discrimination capability. Having the same clothes on a given part of different persons creates a slight variation in the silhouette images, whereas different clothes on the given part of a single subject result in large variations. To overcome this hindrance, we incorporate Fisher discriminant ratios for adaptive weight control. The distribution of z-normalized distances between all pairs of a probe and gallery in each set is calculated and the PDFs obtained. Fisher's discriminant ratio between the PDFs for the same and different subjects is used as the discrimination capability. For example, Fisher's discriminant ratio for the same clothes F^{Sc}_i is defined as

$$F^{Sc}_i=\frac{(\sigma^{Sc}_{B,i})^2}{(\sigma^{Sc}_{T,i})^2} \tag{21}$$



(a) The same clothes (b) Different clothes

Figure 7: PDFs of distances for the same and different subjects

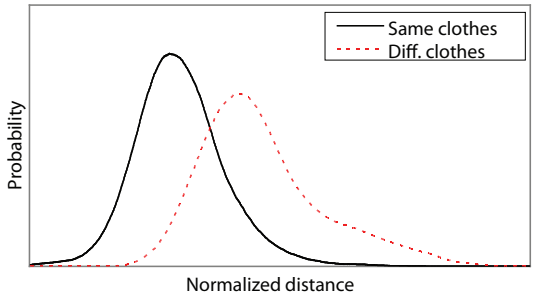


Figure 8: Distance distribution for the same and different clothes

where $(\sigma_{B,i}^{Sc})^2$ and $(\sigma_{T,i}^{Sc})^2$ are the inter-class variance for the distributions of $U_{Ss,i}^{Sc}$ and $U_{Ds,i}^{Sc}$ and the total variance, respectively. The Fisher discriminant ratio for different clothes F_i^{Dc} is calculated similarly.

Figure 7 shows, using the 3rd body part as an example, the PDFs of the z-normalized distances for each set. We can see that the discrimination capability for the same clothes is superior to that for different clothes.

4.5. Probabilities of the same and different clothes

Generally, because the clothing variation exceeds the individual variation, the PDFs of the distances for the same and different clothes have different properties. First, the PDFs of the distances for the same and different clothes are calculated using training sets $U_i^{Sc}(= U_{Ss,i}^{Sc} \cup U_{Ds,i}^{Sc})$ and $U_i^{Dc}(= U_{Ss,i}^{Dc} \cup U_{Ds,i}^{Dc})$, respectively, and the PDFs are obtained by normalizing with the number of pairs in each set (let these be $P_i(d_i|Sc)$ and $P_i(d_i|Dc)$, respectively). As an example, the PDFs for the 3rd body part are illustrated in Fig. 8, which clearly shows the difference between them.

Thus based on the PDF difference, posteriors for the same and different clothes for a given probe are formulated. Let $\boldsymbol{d}_i(\boldsymbol{x}^p; \boldsymbol{X}^G)$ be a set of distances

between $\boldsymbol{x}^p$ and galleries in $\boldsymbol{X}^G$ for the ith body part. The posterior for the same clothes $P_i^{Sc}(\boldsymbol{x}^p; \boldsymbol{X}^G)$ is written according to the Bayesian rule:

$$P_i^{Sc}(\boldsymbol{x}^p; \boldsymbol{X}^G) = \frac{P_i(\boldsymbol{d}_i(\boldsymbol{x}^p; \boldsymbol{X}^G)|Sc)P_i(Sc)}{\sum_{c\in\{Sc,Dc\}} P_i(\boldsymbol{d}_i(\boldsymbol{x}^p; \boldsymbol{X}^G)|c)P_i(c)} \tag{22}$$

$$P_i(\boldsymbol{d}_i(\boldsymbol{x}^p; \boldsymbol{X}^G)|c) = \prod_{\boldsymbol{x}^g \in \boldsymbol{X}^G} P_i(d_i(\boldsymbol{x}^p, \boldsymbol{x}^g)|c), \tag{23}$$

where $P_i(Sc)$ and $P_i(Dc)$ are priors for the same and different clothes, obtained from the ratios of the numbers of pairs in the sets U_i^{Sc} and U_i^{Dc}, respectively.

5. Experiments

5.1. Datasets

We used a subset containing 446 sequences of 20 subjects (10 males and 10 females) from our gait dataset with clothing variations to train the PCA subspace. The sequences consist of different clothing types ranging from 15 to 28 for each subject. These sequences were also used to learn the discrimination capabilities and for the PDFs of the distances for the same and different clothes.

For testing, we used a gallery set from the dataset consisting of the standard clothes sequences of 48 subjects, excluding the 20 training subjects. The probe set included 856 sequences for these 48 subjects with other clothes types, excluding the standard clothes.

The dataset was divided into three sets, a training set (20 subjects with all clothes types), a gallery set (the other 48 subjects with a single clothes type), and a probe set (the same 48 subjects with other clothes types) to separate the training sets and test sets in terms of subjects, and to separate the test gallery and test probe in terms of clothing to ensure strict separation conditions for the experimental evaluations.

5.2. Comparison of matching measures

A subset consisting of 34 subjects, both male and female, from the test dataset was prepared. The clothes types of these subjects were divided into 15 gallery and 15 probe sets. The probe set excludes the corresponding gallery clothes type. The performance is compared using the 4 matching techniques. The comparison results are shown in Fig. 9. It is clear from the experimental results that the average of the 2 minimum distances of the subsequences for a probe and the gallery is the best.

5.3. Contributions of the adaptive weight

We exploit the discrimination capability as a matching weight for each body part and control these weights adaptively. In this section, we show how the adaptive weights are related to the types of clothing. Figure 10 shows the learned Fisher ratios for the same and different clothes for all body parts. We can see

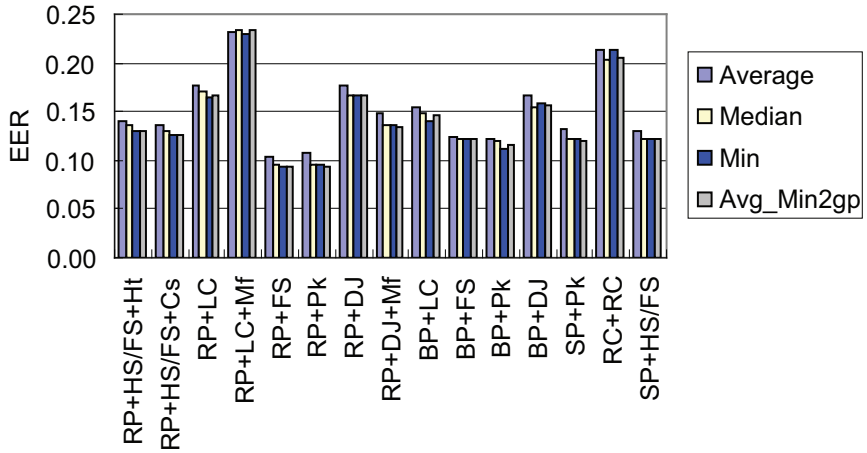


Figure 9: Comparison of the matching measure techniques using different galleries.

that the Fisher ratios for the same clothes are always higher than those for different clothes. Figures 11 and 12 show two typical examples of the calculated posteriors and adaptive weights for clothing types B (RP+DJ) and T (Sk+FS), respectively, in the testing phases of two probes. In both cases, the gallery clothing type is the standard clothes (RP+FS). For clothing type B, only the upper body parts change while the lower body parts are only very slightly affected. The system calculates posteriors and weights (Fig. 11) for each body part and mainly uses the lower parts for effective recognition. Similarly, for clothing type T, only the lower body parts change while the upper body parts are barely affected. The system can adaptively calculate posteriors and weights (Fig. 12) for each body part and mainly uses the upper parts for successful recognition.

5.4. Comparison with other methods

The proposed method was compared with the whole-based [20], LDA [23], SVB frieze pattern [10], gait components-based [14], and CSA+DATER [24] methods to confirm its effectiveness. The whole-based method uses all the frequency-domain features which are dimension-reduced by the PCA in the same way as in the part-based method. The LDA method also uses all the frequency-domain features. First the dimensions are reduced by PCA, and then LDA is applied. The gait component-based method [14] defined 7 components, with the 4 most important components, namely head, arm, trunk and back-leg, being used for identification using average gait image (AGI) features. In CSA+DATER [24], a matrix-based unsupervised algorithm, coupled subspace analysis (CSA), is employed as a preprocessing step to remove noise and retain

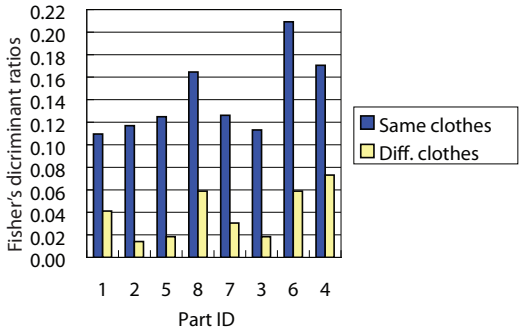
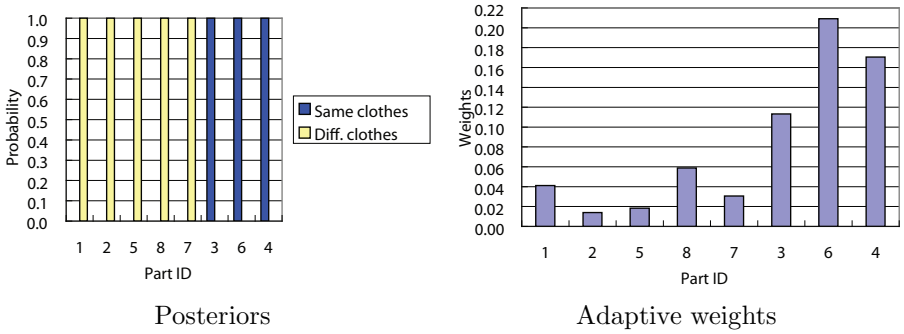


Figure 10: Learned Fisher ratios for the same and different clothes for all body parts. Part ID is sorted from upper to lower body parts.



Posteriors Adaptive weights

Figure 11: Calculated probability and adaptive weights for clothing type B

the most representative information using averaged gait image. Then, a supervised algorithm, discriminant analysis with tensor representation (DATER), is applied to further improve the classification ability. We evaluate the performance of CSA+DATER for various DATER dimensions ranging from 2×2 to 20×20.

Gait identification performances were evaluated by Receiver Operating Characteristics (ROC) curves and the Equal Error Rate (EER) [25]. ROC curves depict the relation between the False Rejection Rate (FRR) and False Acceptance Rate (FAR). The ROC of the CSA+DATER [24] is shown using a similar dimension to LDA. The ROC curves in Fig. 13 show that the part-based method outperforms the other methods. In EER analysis, the lowest EER represents the best recognition performance. We showed the averaged EER with its standard deviation with a range from 2×2 to 20×20 dimensions of DATER for the CSA+DATER [24] method. In Fig. 14, EER decreases significantly for the part-based method compared with the other methods. The results of ROC and

Posteriors Adaptive weights

Figure 12: Calculated probability and adaptive weights for clothing type T

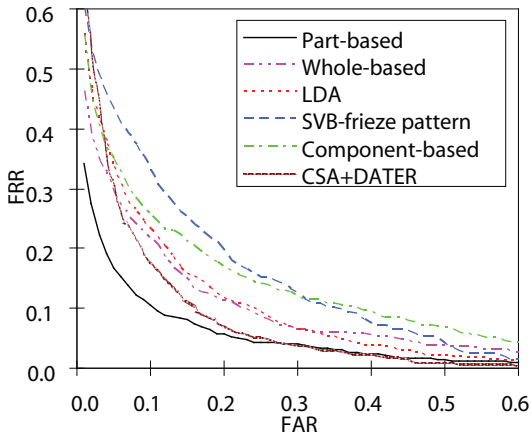


Figure 13: ROC curves for comparison of methods

EER analysis confirm the effectiveness of the proposed method.

Although CSA+DATER is the most competitive to the proposed method, the DATER still has a problem of dimension selection as reported in [24]. On the other hand, the CSA+DATER [24] and the proposed part-based method is not exclusive each other. The DATER can be included as an efficient discriminant analysis subspace for our part-based distance calculation. Therefore, the further investigation of the appropriate discriminant subspace for part-based distance calculation is one of potentials for future works.

5.5. Performance analysis for all combinations of gallery and probe clothing

We prepared a subset consisting of 34 subjects, both male and female, from the test dataset to analyze the combined performance. The subjects were chosen in such a way that the maximum number of common items of clothing could be used as the gallery. The total number of common clothing items in this case

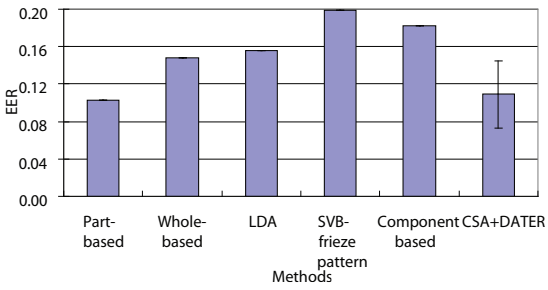


Figure 14: EER with standard deviation for comparison of methods

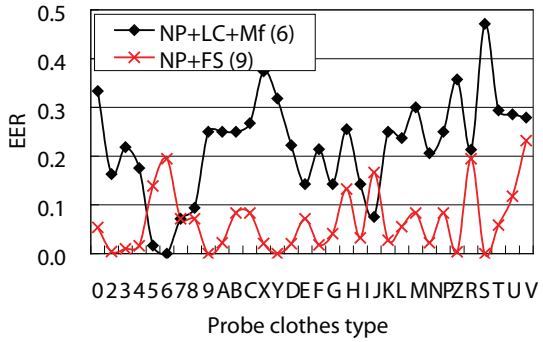


Figure 15: EER of all 32 probe items of clothing for two galleries: RP+LC+Mf and RP+FS.

was 15. The 32 different probe sets were prepared for each gallery clothing type. The number of subjects in the probe sets varied between 2 and 34 depending on the common clothes among the subjects. A recognition test was performed for each probe against all 15 galleries separately and the EER calculated.

Figure 15 shows the EER of two typical galleries (RP+FS and RP+LC+Mf) for all the probe sets. We can see that the average performance of (RP+FS) as the gallery is better than that of (RP+LC+Mf). To choose the best gallery clothes, we analyze the EER for all 15 galleries with a probe set consisting of all the other clothing types. Then, a recognition test is performed separately for each gallery. In Fig. 16, we can see that clothing type (RP+FS) is the best, while clothing type (RP+LC+Mf) is the worst type for the gallery.

Our dataset is biased towards the standard clothes type. In the real world, the clothing types of the general population are also biased in the same way (this is our assumption). The gallery selection results in this paper are derived from the probe prior bias, however this prior does not differ much from that in the real world.

There is a strong relationship between EER and adaptive weight. We have

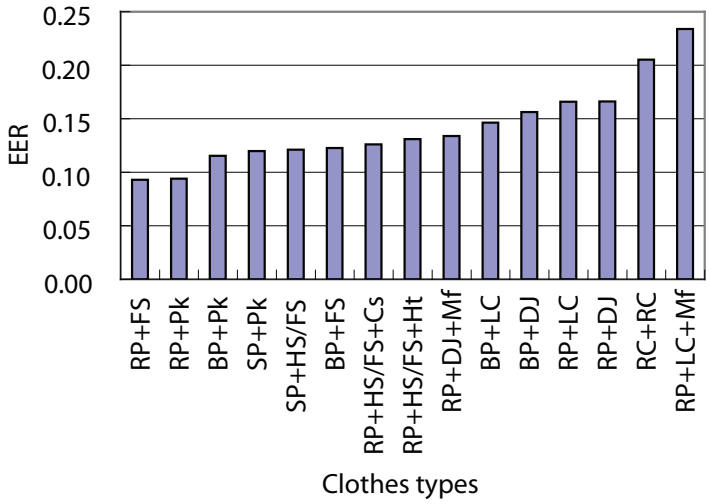


Figure 16: EER for gallery selection

shown the contribution of the weights for each part and how the weights adaptively relate to the clothing types in Section 5.3. We can find the total weight for a given probe and a gallery set as:

$$W(\boldsymbol{x}^p; \boldsymbol{X}^G) = \sum_{i=1}^{N_{parts}} w_i(\boldsymbol{x}^p; \boldsymbol{X}^G) \tag{24}$$

The relationship between the total weight and EER for typical gallery vs. probe examples is depicted in Fig. 17. The total weight increases as the number of similar body parts increases and hence the EER decreases. It is also clear from Fig. 18 that the easy combinations have large body parts with the same clothes, while the difficult combinations have few. Thus the total weight for a given probe is proportional to its recognition accuracy. Figure 19 depicts the relationship between the integrated distances and the total weights.

The above observation can be applied to the following applications:

- Changed-clothing-aware gait identification access control system. The system asks a subject to take off certain clothes based on the probabilities of the body parts when both the total weight and integrated distance are low.

- Adaptive acceptance threshold control for surveillance system. The system can adaptively change the number of candidates to show the guard, which results in low miss detection and diminishes the burden on the guard.

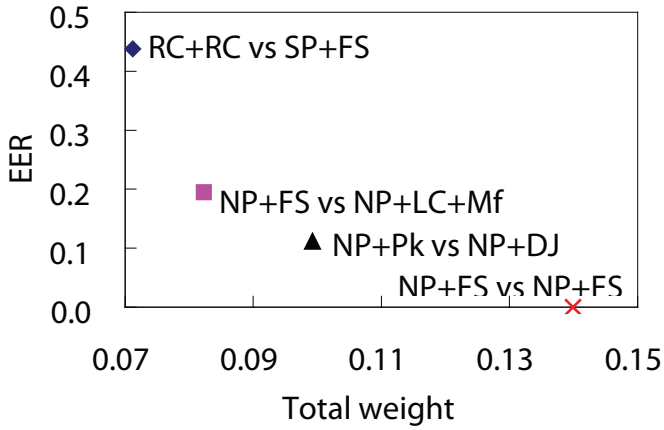


Figure 17: Relationship between total weight and ERR for a typical gallery vs. probe example cases

6. Conclusion and future works

This paper presented a method for part-based gait identification with substantial clothing variations. First, a large-scale gait dataset with clothing variations was constructed. Next, 8 body parts, including 4 overlapping parts, were defined based on anatomical statistics, and dimension-reduced frequency-domain features were used as part-based gait features. Then, during the training phase, discrimination capabilities and PDFs of the distances between a probe and the gallery were acquired separately for the same and different clothes and for each of the 8 parts. Subsequently, given a probe in the test phase, matching weights for the 8 parts were determined adaptively from the distances between the probe and all the galleries based on the trained discrimination capabilities and PDFs. We showed the importance of weights for each part and how the weights adaptively related to the clothing types. The results of the experiments using the constructed dataset, confirm that the proposed method is largely robust in comparison with other methods.

The contributions of the body parts and how to divide the body optimally will be investigated in the future.

References

[1] P. Phillips, S.Sarkar, P. I.Robledo, K. Bowyer, The gait identification challenge problem: Data sets and baseline algorithm, in: Proc. of the 16th Int. Conf. on Pattern Recognition, Vol. 1, Quebec, Canada, 2002, pp. 385–388.

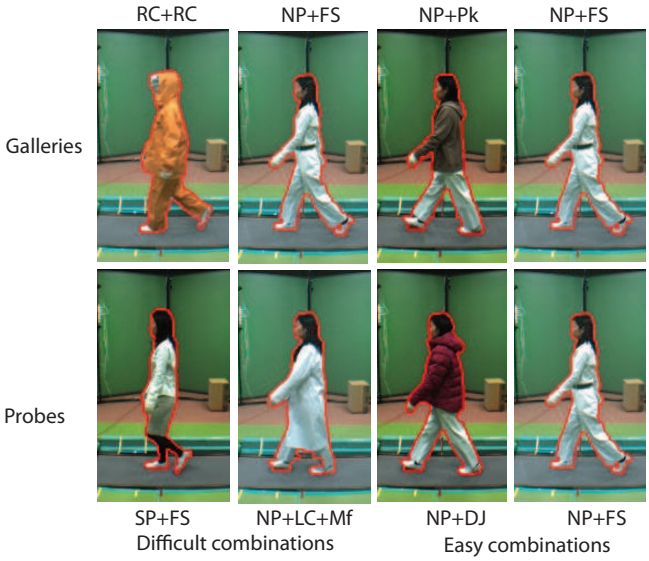


Figure 18: Typical examples of easy and difficult gallery-probe combinations

[2] L. Wang, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1505–1518.

[3] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans usinggait, IEEE Trans. on Image Processing 13 (9) (2004) 1163–1173.

[4] J. E. Boyd, Synchronization of oscillations for machine perception of gaits, Computer Vision and Image Understanding 96 (2004) 35–59.

[5] M. Nixon, J. Carter, Automatic recognition by gait, Proc. of the IEEE 94 (11) (2006) 2013–2024.

[6] Z. Zhou, A. P.-Bennett, R. Damper, A bayesian framework for extracting human gait using strong prior knowledge, IEEE Trans. on Pattern Analysis and Machine Intelligence 28 (11) (2006) 1738–1752.

[7] Z. Liu, S. Sarkar, Effect of silhouette quality on hard problems in gait recognition, Trans. of Systems, Man, and Cybernetics Part B: Cybernetics 35 (2) (2005) 170–183.

[8] C. Lee, A. Elgammal, Carrying object detection using pose preserving dynamic shape models, in: AMDO06, Springer, 2006, pp. 315–325, lNCS 4069.

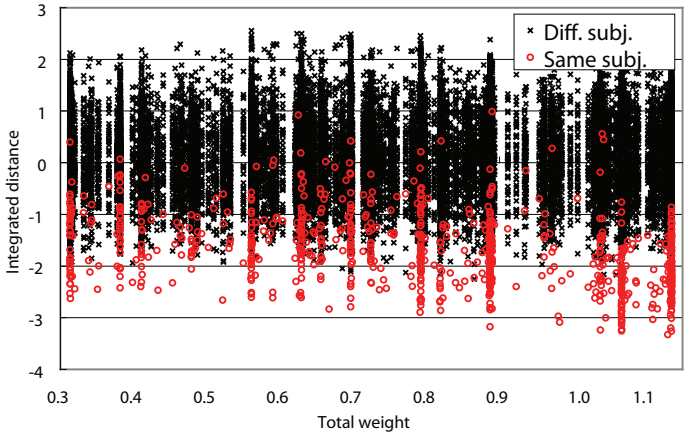


Figure 19: System confidence measure

[9] J. Han, B. Bhanu, Individual recognition using gait energy image, Trans. on Pattern Analysis and Machine Intelligence 28 (2) (2006) 316– 322.

[10] S. Lee, Y. Liu, R. Collins, Shape variation-based frieze pattern for robust gait recognition, in: Proc. of the 2007 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007, pp. 1–8.

[11] C. BenAbdelkader, R. Culter, H. Nanda, L. Davis, Eigengait: Motion-based recognition people using image self-similarity, in: Proc. of Int. Conf. on Audio and Video-based Person Authentication, 2001, pp. 284–294.

[12] L. Lee, W. Grimson, Gait analysis for recognition and classification, in: Proc. of the 5th IEEE Conf. on Face and Gesture Recognition, Vol. 1, 2002, pp. 155–161.

[13] N. Boulgouris, Z. Chi, Human gait recognition based on matching of body components, Pattern Recognition 40 (6) (2007) 1763–1770.

[14] X. Li, S. Maybank, S. Yan, D. Tao, D. Xu, Gait components and their application to gender recognition, IEEE Transactions on Systems, Man, and Cybernetics,Part C 38 (2) (2008) 145–155.

[15] S. X. Li, D. Tao, Gender recognition based on local body motions, 2007, pp. 3881–3886.

[16] J. Shutler, M. Grant, M. Nixon, J. Carter, On a large sequence-based human gait database, in: Proc. of the 4th Int. Conf. on Recent Advances in Soft Computing, Nottingham, UK, 2002, pp. 66–71.

[17] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The humanid gait challenge problem: Data sets, performance, and analysis, Trans. of Pattern Analysis and Machine Intelligence 27 (2) (2005) 162–177.

[18] S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in: Proc. of the 18th Int. Conf. on Pattern Recognition, Vol. 4, Hong Kong, China, 2006, pp. 441–444.

[19] W. Dempster, G. Gaughran, Properties of body segments based on size and weight, American Journal of Anatomy 120 (1) (1967) 33–54.

[20] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in: Proc. of the 9th European Conf. on Computer Vision, Vol. 3, Graz, Austria, 2006, pp. 151–163.

[21] G. W. Brown, On small-sample estimation, The Analysis of Mathematical Statistics 18 (4) (1947) 582–585.

[22] J. H. W. Jr., Hierarchical grouping to optimize an objective funtion, Journal of the American Statistical Association 58 (301) (1963) 236–244.

[23] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, Trans. on Pattern Analysis and Machine Intelligence 28 (2006) 863–876.

[24] D. Xu, S.Yan, D.Tao, L.Zhang, X.Li, H. Zhang, Human gait recognition with matrix representation, EEE Trans. on Circuits and Systems for Video Technnology 16 (7) (2006) 896–903.

[25] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The feret evaluation methodology for face-recognition algorithms, Trans. of Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.