

# Recognizing Night Walkers Based on One Pseudoshape Representation of Gait

Daoliang Tan, Kaiqi Huang, Shiqi Yu, and Tieniu Tan  
National Laboratory of Pattern Recognition, Institute of Automation  
Chinese Academy of Sciences, Beijing, P.R. China, 100080  
{dltan, kqhuang, sqyu, tnt}@nlpr.ia.ac.cn

## Abstract

*Gait is a promising biometric cue which can facilitate the recognition of human beings, particularly when other biometrics are unavailable. Existing work for gait recognition, however, lays more emphasis on the problem of daytime walker recognition and overlooks the significance of walker recognition at night. This paper deals with the problem of recognizing nighttime walkers. We take advantage of infrared gait patterns to accomplish this task: 1) Walker detection is improved using intensity compensation-based background subtraction; 2) pseudoshape-based features are proposed to describe gait patterns; 3) the dimension of gait features is reduced through the principal component analysis (PCA) and linear discriminant analysis (LDA) techniques; 4) temporal cues are exploited in the form of the relevant component analysis (RCA) learning; 5) the nearest neighbor classifier is used to recognize unknown gait. Experimental results justify the effectiveness of our method and show that our method has an encouraging potential for the application in surveillance systems.*

## 1. Introduction

Much emphasis has been placed on the importance of security over the past years, especially after the 9.11 attack. In this context, gait is gaining more and more attention not only from academic researchers, but also from industrial practitioners, though it is still in its infancy. Human gait carries a personal style in which one person moves. The most critical advantage of gait over conventional biometrics (such as face, iris, and fingerprint) lies in its perceivability at a long distance. The drawback of gait resides in its possible change in the motion style after a long-time period due to its behavioral attribute, and this poses a great challenge for robust gait recognition algorithms. The work by Murray *et al.* [15] appeared to give a support for the uniqueness of gait to individuals. Later, Cutting and Kozlowski [6] found out that people have the ability to recognize their friends using motion cues. The evidence inspires researchers in the

computer vision field to leverage gait for automatic walker recognition. It is this huge potential in gait that drives a great deal of work on intelligent gait recognition in an effort to strengthen our security level.

But we can see from a careful review of the past work that there is still a lack of research on gait-based walker recognition at night. In other words, night gait recognition has not been sufficiently studied in the literature. On the other hand, night surveillance should be an indispensable component for an intelligent visual surveillance system. This necessitates the investigation of the night walker recognition issue. In addition, the previous finding naturally reflects the importance and necessity of this work both from the perspective of gait recognition and from the point of view of night surveillance. We use the *CASIA Infrared Night Gait Dataset* [20] to facilitate this research. Moreover, this work is more systematic than Tan *et al.*'s [20] in that we recognize all the subjects in the dataset, rather than a fraction of walkers. First, we preprocess each incoming video frame to reduce as much the effect of halo and dynamic gain on detection results as possible. Then, background subtraction is applied to localize moving people, with fusing motion cues derived from pairwise consecutive frames. Additionally, this paper proposes a gait representation, known as normalized pseudo-height and width (NPHW), to characterize walker signatures. Moreover, we employ informative linear subspace projections to reduce the dimension of raw gait features and combine temporal cues within gait in a machine learning-based form. Finally, the recognition of unknown gait features is accomplished using nearest neighbor-based decision making.

The remainder of this paper is as follows. Section 2 gives a review of related work. Described in Section 3 is our method. Then we justify the effectiveness of our method in Section 4. Finally, Section 5 concludes this paper.

## 2. Related Work

Modern gait recognition stems from the pioneering experiments by Niyogi and Adelson [16]. Prior methods for gait recognition can be generally grouped into two cate-

gories: one is based on structural models and the other on appearance features. Model-based methods are less sensitive to viewpoint variation than appearance-based ones, yet at the price of intensive computation. Choosing whether to use model approaches or appearance ones depends on applications.

In general, model-based work begins with a simplified structural model (2D or 3D) that may incorporate motion constraints. Then, salient features are extracted from video images which can be used to fit parameters of the structural model. Finally, the parameters produce gait signatures in a direct or indirect manner which help with the recognition of unknown walkers. For example, Niyogi and Adelson [16] employed a skeleton model to characterize the structure of human beings and fulfilled the recognition with the aid of four angles in the skeleton model. In addition, the work [5] exploited two connected pendulums to describe the movement of human legs and discriminated individuals by means of the Fourier coefficients within the range of the low frequency band. Later, Urtasun and Fua [22] utilized a vector of angles in a 3D model to differentiate gait patterns. It is worthwhile to point out that as far as human structural models are concerned, there is no general guideline to make a reasonable compromise between generality and specificity.

On the contrary, appearance-based approaches achieve gait recognition through the direct use of conspicuous features in each video frame. For instance, Kale *et al.* [12] made use of the vector of the outermost widths of binary human silhouettes to distinguish gait patterns from unknown walkers, whereas Bobick and Johnson [2] made a distinction between gait patterns based on the static parameters in the human body. In addition, Collins *et al.* [3] took advantage of four-key-frame-based template matching to recognize unknown gait. Furthermore, Han and Bhanu [9] obtained a progressive identification performance using the GEI description. Moreover, Tao *et al.* [21] explored the use of tensor data expression to recognize the carrying gait. In particular, Sarkar *et al.* [18] established a large daytime gait dataset for the gait recognition community which was accompanied by a baseline algorithm. This work made an attempt to provide a benchmark for evaluation of gait recognition algorithms, though its algorithm is somewhat time consuming. However, it is imperative to solve the problem of how to generalize the discriminative power of appearance features across different conditions. As well as that, one of the bottlenecks of current gait recognition algorithms is pedestrian detection with good robustness.

Fortunately, infrared imaging provides a promising prospect for light-sensitive detection algorithms on account of its light irrelevance and shadow immunity, especially with the maturity of the technology in infrared imaging and the decrease of the cost of infrared imaging devices. For example, Davis and Sharma [7] dealt with the problem of

robust pedestrian detection based on univariate Gaussian background models, with special attention paid to thermal halo effect. In addition, Han and Bhanu [10] discussed the problem of recognizing human activity in thermal infrared images. More recently, Yin and Collins [24] built on motion history images to localize moving objects in thermal video sequences. In particular, Tan *et al.* [20] made a preliminary effort to recognize night gait in thermal infrared videos. It should be pointed out that much of prior work for object detection in infrared imagery ignores halo effect, polarity change, and dynamic gains which have a nonnegligible impact on detection results. Nevertheless, these attempts do actually help attract more and more researchers to join the alliance which hopes to detect, track, and recognize objects beyond the visible spectrum.

On the other hand, advances in psychology, neuroimaging, and computer vision present some evidence for the viability of detouring the gait cycle estimation step in the conventional route toward automated gait recognition. First of all, recent studies in experimental psychology found out that visual experience has a significant impact on identity perception [11]. Then, the body-inversion effect [17] in cognitive psychology indicates that the perception of human body may proceed configurally or globally. More surprisingly, a hemianopic patient [4] who suffered from the inability to recognize forms from motion cues could detect motion and discriminate static shape stimulus. Moreover, Downing *et al.* [8] discovered a region in the lateral occipitotemporal cortex which is category specific to the visual perception of appearance of human bodies (apart from face) and is insensitive to motion cues. Furthermore, the work [23] concluded that shape cues play a more critical role in gait recognition than those provided by motion information. These findings inspire us to form a new gait view: It is not motion but appearance (shape) that dominates gait recognition, and each human gait is a stochastic realization of one's stances. Hence, we speculate based on the statistics (visual experience) that for one fixed viewpoint, each binary human silhouette is a random realization for 2D projections of gait sequences. This view induces the notion of equivalence constraints (two samples come from the same class) in machine learning, which makes it natural to incorporate the RCA learning [1] into gait recognition. The benefit of this view is that it can circumvent the step of estimating gait cycles.

### 3. Method

Our method is a bottom-up procedure. This algorithm comprises six modules: preprocessing, detection, tracking, gait representation, subspace projection, and classification. The role of the preprocessing module is in reducing as much the influence of external disturbances on object detection as possible. Then the detection module is responsible for lo-

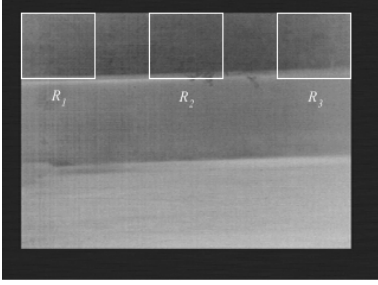


Figure 1. Three regions having no severe thermal variation.

calizing walkers which provides significant cues for people tracking. In addition, the gait representation module serves as extracting simple yet informative gait features. Furthermore, the purpose of subspace projection is to reduce data redundancy and computation load as well as to enhance the separability of data. Finally, the classification module achieves the recognition of unknown gait based on the nearest neighbor decision rule. The following will give a more detailed description of each module in our algorithm.

### 3.1. Preprocessing

A good preprocessing module can greatly facilitate the subsequent detection. But most of prior work pertaining to object detection in the infrared spectrum usually ignored the problem of dynamic gain in thermal infrared imagery which can result in unnecessary intensity fluctuation and noisy (or erroneous) cues in a vicious circle. Our method utilizes intensity compensation scheme to weaken the disturbance of dynamic gain.

Here we assume that there does not exist abrupt changes in thermal radiation over short time durations in some scene regions. This assumption is reasonable, since temperature change must be gradually progressive. First of all, the preprocessing module pinpoints the regions without severe thermal variation over short time intervals. Currently this step is accomplished in an interactive way: We manually specify these regions through mouse clicking in the video window. The reason for the use of the manual mode is to simplify computation in view of the limited computational resources. There are three regions ( $R_1$ ,  $R_2$ , and  $R_3$ ) which are regarded as not having great thermal variation, as shown in Figure 1. At the same time, we also use the background image  $I_B$  as a reference image  $I_r$  which will benchmark compensation and can be updated in a weighted fashion until 20 frames. For our experiments, image averaging can basically produce acceptable background images. A practical surveillance system can attempt to use the multiple Gaussian models (e.g., [19]) to more flexibly characterize the dynamical background. It should be noted that the regions remain the same for all the thermal video sequences. Then, the module applies Gaussian filtering to each incoming im-

---

#### Algorithm 1 Preprocessing

---

1.  $I_t = I_t * K_g$  where the symbols “\*” and “ $K_g$ ” mean the convolution and a Gaussian kernel, respectively.
  2.  $C_{i,t} = \frac{1}{N_i} \sum_{(x,y) \in R_i} I_t(x,y) - I_r(x,y)$  ( $i = 1, 2, \dots, M$ ) where “ $M$ ” and “ $N_i$ ” represent the numbers of the regions without severe thermal variation and the pixels in  $R_i$ , respectively.
  3.  $C_t = \frac{1}{M} \sum_{i=1}^M C_{i,t}$ .
  4.  $I_t(x,y) \leftarrow I_t(x,y) - C_t$ .
- 

age  $I_t$  in an attempt to remove irrelevant random noise. In addition, we calculate an average intensity difference  $C_{i,t}$  for each pinpointed region in between the new image  $I_t$  and the reference image  $I_r$ . Finally, the average of the intensity differences is used to compensate the new image  $I_t$ , as indicated in Step 4 of Algorithm 1.

### 3.2. Detection and Tracking

Object detection plays an important role in intelligent visual systems due to its backbone function at later stages of video analysis. This paper takes advantage of background subtraction-based walker localization similar to the work by Davis and Sharma [7], in combination with motion cues and geometric constraints. But our method differs from [7] in that we employ simple image operation to eliminate halo effect rather than complex (and time-consuming) image processing techniques.

In particular, we exploit the frame difference between two consecutive video frames (after intensity compensation) to define a coarse detection region that can to some degree remove noise. Then we compute an absolute difference image and a relative difference image between the current frame and the background image. Finally, these two difference images are combined to refine the detected region by virtue of image logical operations, in conjunction with geometric constraints in human bodies (distances between body parts). Algorithm 2 gives a description of this procedure. At the same time, Figure 2 illustrates related detection results. It can be seen from Figure 2 that histogram equalization brings much noise into images, in spite of better image contrast, and that our algorithm can produce satisfactory results, at least on this dataset. It is worth pointing out that we can further apply morphological filtering to remove small holes in binary silhouettes. However, unlike the requirement of [12] for silhouettes, in our gait representation, this step is optional. In addition, tracking can be readily completed by nearest neighbor rule because of this simplified case—one walking person in the scenario.

---

**Algorithm 2** Algorithm for walker detection

---

1.  $FD_t(x, y) = |I_t(x, y) - I_{t-1}(x, y)|$ .
  2.  $FD_t(x, y) = \begin{cases} 255 & FD_t(x, y) > T_h \\ 0 & \text{otherwise} \end{cases}$   
where  $T_h$  is a threshold.
  3. Connected component analysis for the image  $FD_t$ .
  4. Find a minimum rectangle  $R_m$  which contains all the motion regions.
  5.  $F_{1,t}(x, y) = |I_t(x, y) - I_{B,t}(x, y)|$  where  $I_{B,t}(x, y)$  is the background image at time  $t$ .
  6.  $F_{1,t}(x, y) = \begin{cases} 255 & F_{1,t}(x, y) > T_{h1} \ \& \ (x, y) \in R_m \\ 0 & \text{otherwise} \end{cases}$   
where  $T_{h1}$  is a threshold.
  7.  $F_{2,t}(x, y) = I_t(x, y) - I_{B,t}(x, y)$  (or  $I_{B,t}(x, y) - I_t(x, y)$ ).
  8.  $F_{2,t}(x, y) = \begin{cases} 255 & F_{2,t}(x, y) > T_{h2} \ \& \ (x, y) \in R_m \\ 0 & \text{otherwise} \end{cases}$   
where  $T_{h2}$  is a threshold.
  9.  $F_{1,t} \cdot F_{2,t} = \begin{cases} F_{1,t} \text{ AND } F_{2,t} & (x, y) \in R_l \text{ (see Fig. 2c)} \\ 0 & \text{otherwise.} \end{cases}$
  10.  $(F_{1,t} + F_{2,t}) = \begin{cases} F_{1,t} \text{ OR } F_{2,t} & (x, y) \in R_h \text{ (see Fig. 2c)} \\ 0 & \text{otherwise.} \end{cases}$
  11.  $F_t = F_{1,t} \cdot F_{2,t} + (F_{1,t} + F_{2,t})$ .
  12. Morphological filtering (optional).
- 

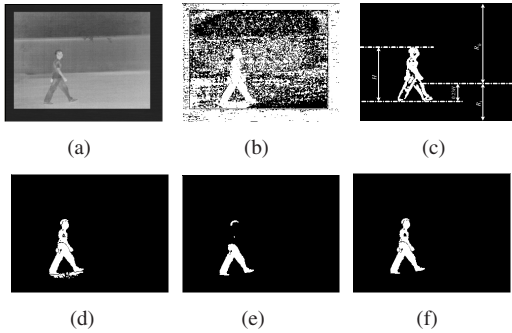


Figure 2. An example for detection. (a) Source frame. (b) Detection result from histogram equalization. (c) Motion Region (Step 2 of Algorithm 2). (d) Result from Step 6 of Algorithm 2. (e) Result from Step 8 of Algorithm 2. (f) Final result using our detection algorithm.

### 3.3. Gait Representation

The purpose of gait representation is to describe the essence of gait patterns and discard the redundancy in raw gait data. In general, this representation has enormous influence upon the final performance of gait recognition systems. A discriminating representation should not only simplify the recognition of gait patterns, but also reduce computa-

---

**Algorithm 3** Gait Representation

---

1.  $H_i = \{(x, i) | F(x, i) \neq 0\}$  ( $i = 1, 2, \dots, L$ ).
  2.  $W_j = \{(j, y) | F(j, y) \neq 0\}$  ( $j = 1, 2, \dots, L$ ).
  3.  $\tilde{H}(i) = \frac{|H_i|}{\max_x |H_i|}$ .
  4.  $\tilde{W}(j) = \frac{|W_j|}{\max_y |W_j|}$ .
  5.  $x = (\tilde{H}^T, \tilde{W}^T)^T$ .
- 

tional complexity. However, methods concerned with feature representation are application specific and have no general guidelines. Fortunately, recent gait research revealed that shape cues play a primary role in gait recognition [23]. This finding enlightens us to extract shape features for gait identification.

More specifically, we first normalize human silhouette images to the same size of  $32 \times 32$ . Then, as is shown in Figure 3, each size-normalized silhouette image is projected in the horizontal and vertical directions, respectively. Meanwhile, we can obtain for each projection direction a maximum integer value which counts the maximal number of foreground pixels in that direction. In addition, the two directional projections are further standardized using the two maximums. Lastly, we make use of the concatenation of the two projections to describe the human gait pattern in the current frame. For sake of convenience, this gait representation is referred to as normalized pseudo-height and width (NPHW). The whole process is illustrated in Algorithm 3. It is worthwhile to point out the relationships between NPHW and another two gait features closely related to the NPHW. On the one hand, the NPHW differs from Liu *et al.*'s frieze patterns [13] in that Liu *et al.* separately utilize the two unstandardized projections to characterize gait, but we employ correlated and standardized patterns to represent human gait. On the other hand, the difference between the representation of Kale *et al.* [12] and the NPHW lies in that the projections used in this paper are two absolute (or pure) counter of foreground pixels, whereas Kale *et al.* use a relative coarse counter for foreground pixels without taking into account the vertical cues. As far as the reconstruction of human shape is concerned, the NPHW representation is merely a partial description for human shape. That is the reason for calling it pseudoshape. Figure 3 gives an illustration for the NPHW gait representation.

### 3.4. Subspace Projection

Subspace projection often performs the role of dimension reduction due to the curse of dimensionality. In a nutshell, the purpose of subspace projection is three-fold: 1) to reduce data dimension; 2) to remove data correlation; and 3) to enhance data separability. In this paper, we expect to



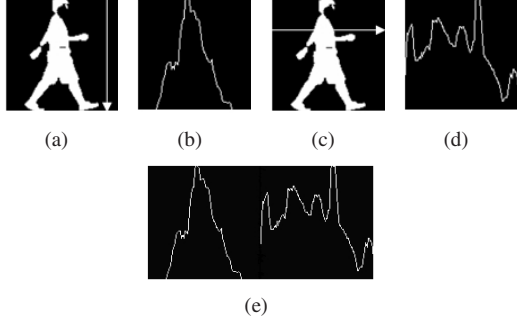


Figure 3. An illustration of the gait representation used in this paper. (a) Direction for the vertical projection. (b) Curve from (a). (c) Direction for the horizontal projection. (d) Curve from (c). (e) Concatenation of (b) and (d).

use linear discriminant analysis (LDA) to improve the feature separability. The singularity issue, however, prohibits the direct use of LDA to transform feature data. One solution is to precede LDA with principal component analysis (PCA).

The PCA projection matrix  $P_{pca}$  can be derived from minimizing the reconstruction error in the mean-square-error sense. This can be formulated as the following optimization problem:

$$\begin{aligned} \min_{P_{pca} \in \mathcal{R}^{s \times d}} \quad & E \| P_{pca} y - x \|^2 \\ \text{s.t.} \quad & P_{pca}^T P_{pca} = I \end{aligned} \quad (1)$$

where  $y \in \mathcal{R}^d$  is the feature vector projected from  $x \in \mathcal{R}^s$ , i.e.,  $y = P_{pca}^T x$ —suppose that the data has been centralized. The constraint  $P_{pca}^T P_{pca} = I$  reinforces the geometric topology that the basis of the lower-dimensional subspace be orthonormal. The problem (1) can be further reduced to Problem (2) using Lagrange multipliers

$$\min_{P \in \mathcal{R}^{s \times d}} f(P, \Lambda) = -tr P^T C_x P + tr \Lambda (P^T P - I) \quad (2)$$

where  $C_x$  is the covariance matrix of  $x$ , and  $\Lambda$  is a symmetric matrix composed of Lagrange multipliers. Making the gradient of  $f(P, \Lambda)$  with respect to  $P$  be zero yields

$$C_x P = P \Lambda \quad (3)$$

Now it is easy to prove that  $P_{pca}$  boils down to the matrix whose columns are the  $d$  eigenvectors of  $C_x$  corresponding to the first  $d$  largest eigenvalues of  $C_x$ . One simple yet useful rule for the choice of  $d$  is as follows [14]:

$$d = \min_{k \in \Phi_e} k \quad (4)$$

$$\Phi_e = \left\{ k \left| \frac{\sum_{i=1}^k \lambda_i}{tr(C_x)} \geq T_d, C_x p_i = \lambda_i p_i, \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s \right. \right\}$$

where  $T_d$  is a threshold within the range 0.95–0.98, and  $p_i$  is the eigenvector of  $C_x$  which corresponds to the  $i$ -th largest eigenvalue  $\lambda_i$  of  $C_x$ . In this paper, we use  $T_d = 0.95$ . Nevertheless, only in the sense of minimizing reconstruction error is  $P_{pca}$  an optimal projection. In fact,  $P_{pca}$  takes not into consideration how to make samples in the same class as compact as possible and those in the different classes as apart as possible.

In this context, the LDA criterion gives an explicit incorporation of the two concerns into its objective function:

$$\max_{P_{lda}} \frac{tr(P_{lda}^T S_B^Y P_{lda})}{tr(P_{lda}^T S_W^Y P_{lda})} \quad (5)$$

where  $S_B^Y$  and  $S_W^Y$  are inter-class scatter matrix and intra-class scatter matrix, respectively. After introducing Lagrange multipliers once again, we can simplify the problem (5) as (6):

$$\min_P g(P, \Gamma) = -tr P^T S_B^Y P + tr \Gamma (P^T S_W^Y P - I) \quad (6)$$

where  $\Gamma$  is a diagonal matrix. Evaluating the derivative of  $g(P, \Gamma)$  with respect to  $P$  and letting it be zero bring the necessary condition (7) that the optimal  $P_{lda}$  must satisfy:

$$S_B^Y P = S_W^Y P \Gamma. \quad (7)$$

This is a typical generalized eigenvalue problem and can be efficiently solved using the matrix SVD. At the same time, this paper just applies the proportion 98% of the total generalized “variance” analogous to the choice of  $d$  in an effort to further reduce the dimension of gait features.

Furthermore, the proposed view involving gait recognition naturally elicits the notion of equivalence constraints—two consecutive gait feature vectors come from the same class. The constraints imply the temporal cues among gait, despite their coarseness. In order to make good use of this clue, we attempt to learn another linear transformation  $P_{rca}$  ( $z = P_{rca} y$ ) or a mathematically equivalent Mahalanobis metric  $P_{rca} P_{rca}^T$  on the subspace spanned by the columns of  $P_{pca} P_{lda}$ . The relevant component analysis (RCA) algorithm [1] actually suits this expectation. In fact,  $P_{rca}$  can be obtained by solving the problem (8):

$$\max_P \log |P| \text{ s.t. } \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \| z_j^i - m_i \|^2 \leq 1 \quad (8)$$

where  $n$ ,  $c$ , and  $n_i$  are the numbers of total samples, classes, and samples within the  $i$ -th class, respectively, and  $m_i$  is the mean of the samples in the  $i$ -th class. After applying the similar matrix technique to (2) and (6) and discarding the scale factor, this paper uses the  $P_{rca}$  with the form (9):

$$P_{rca} = (S_W^Y)^{-\frac{1}{2}}. \quad (9)$$

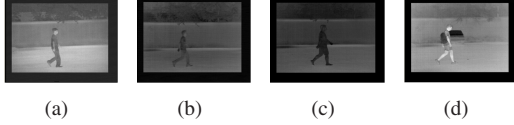


Figure 4. Sample night gait images in the thermal infrared spectrum. (a) Normal walking. (b) Fast walking. (c) Slow walking. (d) Normal walking with a bag.

The optimality of the  $P_{rca}$  in (9) lies in two aspects: one is that it lowers the intra-class variability by the degree to which the within-class distance is minimized; the other is that it maximizes the mutual information of data before and after its transformation,  $z = P_{rca}y$ , under the constraint that  $P_{rca}$  is an invertible linear transformation [1]. As a result, we can readily obtain in the light of the aforementioned subprojections a composition projection matrix  $P_f$  which directly projects  $x$  in the raw feature vector space  $\mathbf{X}$  to  $z$  in the dimension-reduced space  $\mathbf{Z}$ :

$$P_f = P_{rca}P_{lda}^T P_{pca}^T \quad (10)$$

$$z = P_f x. \quad (11)$$

### 3.5. Classification

This paper takes advantage of the simple nearest neighbor classifier to facilitate the recognition of an unknown gait sequence  $U$ . Let  $x_i^u (i = 1, 2, \dots, N)$  be the  $i$ -th raw feature vector in  $U$  and  $E_j (j = 1, 2, \dots, c)$  be an exemplar of the  $j$ -th subject in the low-dimensional space  $\mathbf{Z}$ . We simplify the description of the unknown gait sequence  $U$  by reducing  $U$  to the mean  $\bar{x}^u$  of feature vectors in  $U$ , as shown in (12). Then, we project  $\bar{x}^u$  to  $\bar{z}^u$  expressed by (13). Furthermore, as is displayed in (14), this paper bases the similarity measure on the Euclidean distance in the space  $\mathbf{Z}$ . The whole process is described in (12)–(15):

$$\bar{x}^u = \frac{1}{N} \sum_{i=1}^N x_i^u \quad (12)$$

$$\bar{z}^u = P_f \bar{x}^u \quad (13)$$

$$d_j = (\bar{z}^u - E_j)^T (\bar{z}^u - E_j) \quad (14)$$

$$l(U) = \arg \min_j d_j. \quad (15)$$

## 4. Experiments

For the purpose of facilitation of the research of night walker recognition, Tan *et al.* [20] established an ad hoc night gait dataset called *CASIA Infrared Night Gait Dataset*, which is the largest night gait database in terms of the number of subjects. This dataset consists of night gait sequences of 153 subjects (130 males and 23 females) and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with

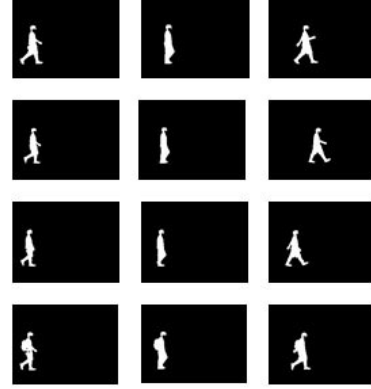


Figure 5. Sample silhouettes from the *CASIA Infrared Night Gait Dataset*. 1st row: normal walking; 2nd row: fast walking; 3rd row: slow walking; 4th row: normal walking with a bag.

a bag. Figure 4 displays four sample images excerpted from this dataset. The sample images present an illustration of the varieties of image quality, with regard to intensity variation, dynamic gain, and halo effect. All these factors pose a great challenge for a night gait recognition algorithm. Nonetheless, this dataset indeed offers an important and valuable complement from the perspective of the diversity of gait data.

### 4.1. Results

We perform our experiments on the basis of the *CASIA Infrared Night Gait Dataset*. Figure 5 depicts sample silhouette images obtained by our detection algorithm. According to the specification of Tan *et al.* [20], this paper conducted a total of four experiments in consideration of the four walking conditions. Here we made use of the set of three normal-walking sequences from each subject as the training set and the remaining sequences as the testing set. In addition, we repeated the experiments four times (Each time three different normal walking sequences from each person were chosen as training data) and employed the average of recognition results to report the final performance, since there are four sequences of normal walking gait for each subject. The reason for not using leave-one-out here is that the large number of gait sequences in this dataset would require time-consuming computation if we did. Table 1 presents the details of the experiments. Exp. A is focused on the performance of algorithms under the normal walking condition. Exps. B and C aim to evaluate the impact of walking paces (fast and slow) on recognition performance. Exp. D is motivated to assess how appearance changes affect recognition accuracy.

In addition, PCA and LDA are so widely used in the gait recognition community that we need to make a fair comparison among PCA, PCA+LDA, and PCA+LDA+RCA based on the NPHW gait representation in the hope of justifica-

Exp.	Gallery	Probe	#Gallery Seq.	#Probe Seq.
A	Normal	Normal	459	153
B	Normal	Fast	459	306
C	Normal	Slow	459	306
D	Normal	Bag	459	306

Table 1. Four experiments on the *CASIA Infrared Night Gait Dataset*.

Exp.	PCA	PCA+LDA	PCA+LDA+RCA
A	98.4%	98.0%	98.4%
B	91.8%	93.0%	93.7%
C	82.4%	90.6%	91.3%
D	24.4%	24.2%	24.7%

Table 2. Correct classification rate (rank 1) using the NPHW on the *CASIA Infrared Night Gait Dataset*

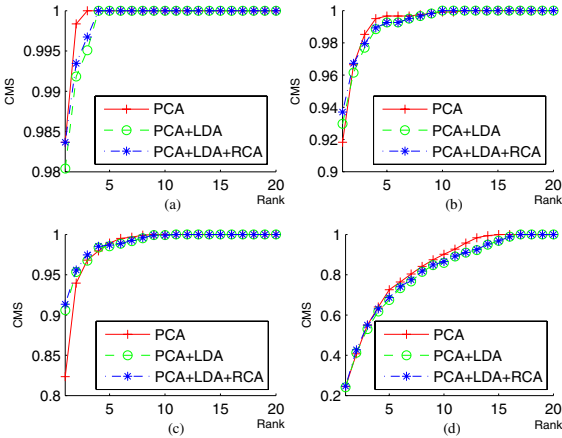


Figure 6. CMS of the four experiments. (a) CMS of normal walking. (b) CMS of fast walking. (c) CMS of slow walking. (d) CMS of normal walking with a bag.

tion of incorporating the RCA learning. To our knowledge, there are few formal algorithmic comparisons at the level of subspace projection in the gait recognition.

Table 2 lists correct classification rates (rank 1) of the four experiments, and Figure 6 shows recognition performance measured by cumulative match score (CMS) for ranks up to 20. Here the CMS at rank  $k$  means the proportion of queries occupying one of the first  $k$  closest resemblances to the real identity. From Table 2 and Figure 6, we can draw the following conclusions:

- The normal walking condition can achieve the best performance, thanks to the similarity between training data and testing data in the aspect of walking traits. Walking pace variation can induce the decline in correct recognition rates which is attributed to mean shape changes caused by varying stride frequency. In our

eyes, the reason for the superiority of fast walking performances over the ones in slow walking is that the variation in stride frequency can lead to the deviation from the mean shape in normal walking, and slow walking has the higher chance of accumulating this error (compared with fast walking). Furthermore, the appearance change has greater impact on the recognition performance, in comparison with other factors—normal walking, fast walking, and slow walking.

- The added RCA learning algorithm can slightly help improve the data separability, as shown in the fourth column of Table 2. This can be explained from a geometric view.  $P_{rca}$  can minimize the within-class distance in the final projection subspace, though not considering between-class cues due to its original clustering purpose. However, we conjecture that  $P_{lda}$  has optimized the between-class topology and that RCA keeps this topology as intact as possible and at the same time shrink the volume of each class. Hence the data separability is consolidated.
- The combination of PCA+LDA+RCA generally outperforms PCA and PCA+LDA in terms of correct classification rate. This is due to the fact that the addition of the RCA projection can further improve the intra-class compactness, which favors classification when using the nearest neighbor rule.
- The LDA projection  $P_{lda}$  preceded by PCA can not always improve correct classification rates over the single  $P_{pca}$ . Its loss in Exp. A and Exp. D is probably due to the fact that the LDA projection misses some useful features for lower dimension which have no active effect on fast walking gait and slow walking gait; however, the RCA projection  $P_{rca}$  can compensate for this loss in the form of better data separability.
- Last but not least, our algorithm is promising, at least in the case of having no great appearance changes, in terms of identification performance and the number of subjects. Although our method is relatively sensitive to great appearance changes, it can still assist in other biometrics (such as face), if available, to reduce the scope of pattern matching in view of its CMS values.

## 4.2. Discussions

The key to the simplicity of our detection algorithm is the assumption on thermal imagery in the scene. If the assumption is violated (e.g., due to the severe self-regulation in the thermal infrared camera), then our detection is likely to introduce some redundant noise. However, the detection algorithm seems to cope well with the self-regulation of moderate level. The degree to which there occurs self-regulation

in good-quality thermal cameras is small and hence the self-regulation is not a vital issue for our method.

Our method avoids the traditional route for automated gait recognition, but it actually makes loose use of temporal cues in the form of equivalence constraints, rather than establishing an HMM model to simulate the temporal relation. Its benefit is in bringing simplified computation, compared with the HMM model. A much subtler scheme is to combine stance clustering with our algorithm for the purpose of making finer distinction on a posture scale.

## 5. Conclusions

This paper has dealt with the problem of night walker recognition with the help of infrared gait patterns. Experimental results on the *CASIA Infrared Night Gait Dataset* [20] indicate that the proposed algorithm makes a promising advance in night gait recognition and that the gait view used is reasonable to some extent, just from the perspective of recognition.

Our contributions include the following: 1) A histogram-based gait representation can be used to characterize walker signatures; 2) this paper makes full use of the notion of equivalence constraints to improve data separability by incorporating the RCA learning into the process of gait recognition; 3) this paper proposes a simple yet useful method for weakening intensity perturbation induced by dynamic gain and halo effect in thermal video which are often ignored by prior work.

## Acknowledgments

This work is supported by the National Basic Research Program of China (2004CB318110), the National Science Foundation of China (60605014, 60332010, 60335010, and 2004DFA06900), and the CASIA Innovation Fund for Young Scientists. Daoliang Tan would like to acknowledge the help of Zhang Zhang and Liangsheng Wang during the collection of the night gait. The authors also thank the anonymous reviewers for their valuable comments and those who contributed to the creation of the CASIA Night Gait Dataset.

## References

- [1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 6:937–965, 2005. 2, 5, 6
- [2] A. Bobick and A. Johnson. Gait recognition using static, activity-specific parameters. In *Proc. CVPR*, 2001. 2
- [3] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *Proc. Automatic Face and Gesture Recognition*, 2002. 2
- [4] A. Cowey and L. M. Vaina. Blindness to form from motion despite intact static form perception and motion detection. *Neuropsychologia*, 38(5):566–578, 2000. 2
- [5] D. Cunado, M. Nixon, and J. Carter. Automatic extraction and description of human gait model for recognition purposes. *CVIU*, 90(1):1–41, 2003. 2
- [6] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5):353–356, 1977. 1
- [7] J. Davis and V. Sharma. Background-subtraction in thermal imagery using contour saliency. *IJCV*, 71(2):161–181, 2007. 2, 3
- [8] P. E. Downing, Y. Jiang, M. Shuman, and N. Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001. 2
- [9] J. Han and B. Bhanu. Statistical feature fusion for gait-based human recognition. In *Proc. CVPR*, 2004. 2
- [10] J. Han and B. Bhanu. Human activity recognition in thermal infrared imagery. In *Proc. CVPR*, 2005. 2
- [11] A. Jacobs and J. Pinto. Experience, context and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5):822–835, 2004. 2
- [12] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. RoyChowdhury, and V. Krueger. Identification of humans using gait. *IEEE Trans. Image Processing*, 13(9):1163–1173, 2004. 2, 3, 4
- [13] Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *Proc. ECCV*, 2002. 4
- [14] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*, pages 242–243. Academic Press, 1979. 5
- [15] M. Murray, A. Drought, and R. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46-A(2):335–360, 1964. 1
- [16] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. CVPR*, 1994. 1, 2
- [17] C. L. Reed, V. E. Stone, S. Bozova, and J. Tanaka. The body-inversion effect. *Psychological Science*, 14(4):302–308, 2003. 2
- [18] S. Sarkar, P. Philips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The human gait challenge problem: data sets, performance and analysis. *PAMI*, 27(2):162–177, 2005. 2
- [19] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR*, 1999. 3
- [20] D. Tan, K. Huang, S. Yu, and T. Tan. Efficient night gait recognition based on template matching. In *Proc. ICPR*, 2006. 1, 2, 6, 8
- [21] D. Tao, X. Li, X. Wu, and S. J. Maybank. Human carrying status in visual surveillance. In *Proc. CVPR*, 2006. 2
- [22] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In *Proc. Automatic Face and Gesture Recognition*, 2004. 2
- [23] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *Proc. CVPR*, 2004. 2, 4
- [24] Z. Yin and R. Collins. Moving object localization in thermal imagery by forward-backward mhi. In *Proc. CVPR*, 2006. 2