


---

Xu, Dong; Yan, Shuicheng; Tao, Dacheng; Zhang, Lei; Li, Xuelong and Zhang, Hong-Jiang (2006). Human gait recognition with matrix representation. *IEEE Transactions on Circuits and Systems for Video Technology* **16** (7) 896-903.

---



---

# Human Gait Recognition With Matrix Representation

Dong Xu, Shuicheng Yan, Dacheng Tao, Lei Zhang, Xuelong Li, and Hong-Jiang Zhang

***Abstract*—Human gait is an important biometric feature. It can be perceived from a great distance and has recently attracted greater attention in video-surveillance-related applications, such as closed-circuit television. We explore gait recognition based on a matrix representation in this paper. First, binary silhouettes over one gait cycle are averaged. As a result, each gait video sequence, containing a number of gait cycles, is represented by a series of gray-level averaged images. Then, a matrix-based unsupervised algorithm, namely coupled subspace analysis (CSA), is employed as a preprocessing step to remove noise and retain the most representative information. Finally, a supervised algorithm, namely discriminant analysis with tensor representation, is applied to further improve classification ability. This matrix-based scheme demonstrates a much better gait recognition performance than state-of-the-art algorithms on the standard USF HumanID Gait database.**

***Index Terms*—Coupled subspaces analysis (CSA), dimensionality reduction, discriminant analysis with tensor representation (DATER), human gait recognition, object representation.**

## I. Introduction

HUMAN gait recognition has attracted growing attention in video-surveillance-based applications [13], such as closed-circuit television (CCTV) surveillance, owing to its great potential in the recognition of individuals from a distance. Recent research [13], [20] has shown that individuals have distinctive and special ways of walking and that human gait recognition has many advantages. First, human gait is a biometric feature that may be captured from a great distance, and, unlike other traditional biometric features, such as fingerprints and irises, gait has the advantage of being unobtrusive.

Numerous schemes [2]–[15], [17]–[21] have been proposed for human gait recognition. They can be roughly divided into two categories: model-based and motion-based approaches. In model-based approaches [3], [5], [14], the human body structure is represented using model parameters fitted based on extracted image features. Motion-based approaches [4], [6]–[8], [10]–[13], [19], [20] employ a compact representation to characterize the motion patterns of the human body without the consideration of the underlying model structure. Our work is motion-based and motivated by the following observations.

- Subspace learning-based algorithms [1], [16] have been successfully employed in face recognition for enhancing recognition accuracy. However, only a few subspace learning schemes [2], [6], [19]–[21] have been applied to gait recognition. Moreover, subspace learning algorithms used in previous research [2], [6], [19]–[21] have been based on the classical principal component analysis (PCA) [16] and linear discriminant analysis (LDA) [1], which consider features only as vectors. It is desirable to investigate recently proposed matrix-based learning algorithms for the gait recognition problem.
- The work in [6] and [11] demonstrates that the gray-level average silhouette over a gait cycle, namely the gait energy image (GEI), is an efficient and effective representation for gait recognition as compared with the time-series-based silhouette representation [13]. An average silhouette is directly in the form of a matrix, and recently we have proposed several dimensionality reduction algorithms [22]–[24] based on matrix representations for face recognition. We also note that two similar works [25], [26] were proposed by Ye *et al.* for face recognition. Better experimental results from matrix-based algorithms have been reported over the traditional PCA [16] and LDA [1] algorithms, especially when the training sets are small.

In this paper, we investigate two subspace analysis algorithms based on matrix representations, namely coupled subspace analysis (CSA) [22], [23] and discriminant analysis with tensor representation (DATER) [24], in the application of gait recognition.

In Section II, we briefly review the human gait database and related gait recognition algorithms. Then, in Section III, the matrix-representation-based learning algorithms CSA and DATER are applied for gait recognition. Experimental results on the standard USF HumanID Gait database [13] are reported in Section IV. Section V states our conclusions.

## II. Brief Review of Gait Database and Related Algorithms

Recently, Sarkar *et al.* [13] constructed the reportedly largest gait database: USF HumanID. This database consists of people walking in elliptical paths while being filmed by two cameras. There are up to five covariates for each individual, namely: 1) viewpoints (left/right); 2) shoe types (A/B); 3) surface types (grass/concrete); 4) carrying status (with/without a briefcase); and 5) time of year (i.e., May/November, the time covariate implicitly contains changes of shoes and clothing). Therefore, there are up to 32 sequences for each person, and a full dataset consists of 1870 sequences from 122 individuals. Sarkar *et al.*

Manuscript received August 19, 2005; revised January 23, 2005. This paper was recommended by Associate Editor I. Ahmad.

D. Xu is with Microsoft Key Laboratory of Multimedia Computing and the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China, and also with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: dongxu@ee.columbia.edu).

S. Yan is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, and also with the Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: scyan@ifp.uiuc.edu).

D. Tao and X. Li are with the School of Computer Science and Information Systems, University of London, Birbieck College, London WC1E 7HX, U.K. (e-mail: dacheng@dcs.bbk.ac.uk; xuelong@dcs.bbk.ac.uk).

L. Zhang and H.-J. Zhang are Microsoft Research Asia, Beijing 100080, China (e-mail: leizhang@microsoft.com; hjzhang@microsoft.com).

Digital Object Identifier 10.1109/TCSVT.2006.877418

TABLE I
USF HumanID GAIT RECOGNITION EXPERIMENTS (V—VIEW; S—SHOE; U—SURFACE; B—BRIEFCASE; T—TIME)

| Probe Set | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Size of Probe Set | 122 | 54 | 54 | 121 | 60 | 121 | 60 | 120 | 60 | 120 | 33 | 33 |
| Gallery/Probe Difference | V | S | VS | U | SU | SV | SUV | B | SB | VB | T | TU |

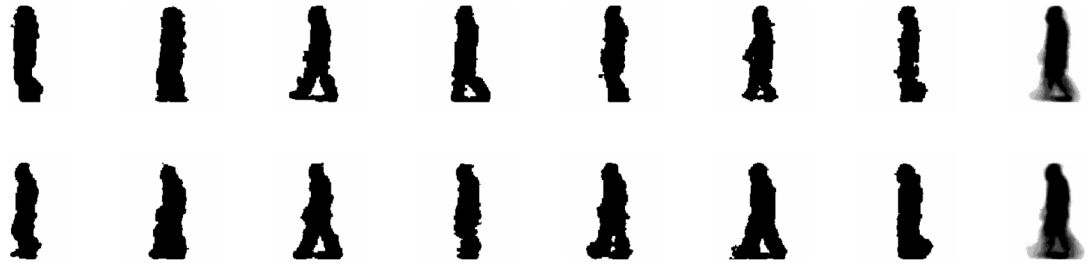

Fig. 1. Normalized and aligned binary silhouettes in the leftmost seven columns, where each row represents a different image sequence of the same person. The rightmost column displays the corresponding gray-level average silhouettes.

fixed one gallery set and created 12 probe sets to test performance under different conditions. Individuals are unique in the gallery and probe sets, and there are no overlapping sequences between the gallery set and probe sets. The differences between the gallery and probe sets are listed in Table I, and more details about the database can be found in [13]. Binary silhouettes are of size 128 by 88 pixels, and some samples are shown in Fig. 1. In [13], a baseline approach is proposed to extract binary human silhouettes and recognize people in the database. Their subsequent work [11] conducted gait recognition based on a similarity comparison of gray-level average silhouettes over different gait cycles. They demonstrated comparable performances with much faster computational speeds.

With the same set of binary silhouettes and the gait periods provided in [13], Han and Bhanu [6] developed a two-stage PCA+LDA algorithm on the so-called GEI for gait recognition. GEI is the gray-level average silhouettes as in [11]. The efficiency of the PCA+LDA strategy has been demonstrated in face recognition [1], in which PCA aims to retain the most representative information and suppress noise for object representation, while LDA aims to pursue a set of features that can best distinguish different objects.

For ease of understanding, we denote the training samples, such as the gray-level average silhouettes in [6] and [11], as $X_i \in \mathbb{R}^{m\times n}$, $i = 1,\ldots,N$, where $N$ is the total number of samples. The class label of each sample $X_i$ is denoted as $l_i$ and the total class number as $N_c$. The samples are assumed to be centered at zero, i.e., $\sum_{i=1}^{N} X_i = 0_{m\times n}$. In [6], each sample $X_i$ is first converted into a one-dimensional (1-D) vector $x_i \in \mathbb{R}^{mn}$. Let $y_i = W_p' x_i$ be the low-dimensional representation from PCA and $x_i^r = W_p W_p' x_i$ be the reconstructed representation; PCA aims to provide the vector-based optimal reconstruction in the least-squares sense according to

$$W_p = \arg\min_{W_p} \sum_i \left\| W_p W_p' x_i - x_i \right\|^2 \tag{1}$$

where $W_p$ is the projection matrix. With $W_p$, each sample $x_i$ is converted to its low-dimensional representation $y_i$. Then, LDA is applied to maximize the between-class scatter and, at the same time, minimize the within-class scatter

$$W_l = \arg\max_{W_l} \frac{\sum_{c=1}^{N_c} n_c \left\| W_l' \overline{y}_c - W_l' \overline{y} \right\|^2}{\sum_{i=1}^{N} \left\| W_l' y_i - W_l' \overline{y}_{l_i} \right\|^2} \tag{2}$$

where $W_l$ is the projection matrix, $\overline{y}_c$ is the average vector of the samples in class $c$, $\overline{y}$ is the total average vector for all the samples, and $n_c$ is the sample number of the $c$-th class. The final projection matrix is obtained by $W = W_p W_l$. With $W$, the samples in the gallery and probe sets can be converted into the low-dimensional representation, and then the nearest neighborhood classifier is used for final classification. For further details, readers are referred to [6].

## III. MATRIX REPRESENTATION FOR GAIT RECOGNITION

Here, we first review our recently proposed matrix-based schemes, CSA [22], [23] and DATER [24], and briefly analyze their learnability and computational complexity. Then, we propose a two-stage scheme, called CSA+DATER, for gait recognition based on gray-level average silhouettes over a gait cycle. In our notation, we define the inner product of two matrices $A$ and $B \in \mathbb{R}^{m\times n}$ as $\langle A, B\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}$ and the Frobenius norm of the matrix $A$ as $\|A\|_F = \sqrt{\langle A, A\rangle}$.

### A. Brief Review of CSA and DATER

In the conventional PCA and LDA algorithms [1], [6], [16], the image matrix is concatenated into a vector, thus the image object is often represented in a very high-dimensional feature space, whereas, in many recognition applications, the number of available training samples is small, typically resulting in the well-known curse of dimensionality and the small sample size problem. Furthermore, in real applications, the extracted feature of an object often has some specialized structure in the form of a second-order or even higher order tensor. For example, the gray-level average silhouette is a second-order tensor or matrix.

Therefore, it would be highly desirable to uncover such underlying structure in gait recognition.

To utilize spatial structure information and overcome the curse of dimensionality as well as small sample size problems, we proposed an unsupervised algorithm, namely CSA [22], [23], in our previous research. We also used a supervised algorithm, namely DATER [24], to conduct dimensionality reduction on the objects represented as matrices or higher order tensors. In the following, we take the second-order tensor (matrix) representation as an example to review the CSA and DATER algorithms.

*1) CSA:* Denote $U \in \mathbb{R}^{m' \times n'}$ and $V \in \mathbb{R}^{n \times n'}$ as two projection matrices to pursue, with $Y_i = U'X_iV \in \mathbb{R}^{m' \times n'}$ as the low-dimensional matrix representation for sample $X_i$ and $X_i^r = UU'X_iVV'$ as the reconstructed matrix. In the dimensionality reduction task, usually we have $m' \leq m$ and $n' \leq n$. As shown in [22] and [23], CSA aims at the optimal reconstruction directly based on the matrix representation

$$(U^*, V^*) = \arg\min_{U,V} \sum_i \|UU'X_iVV' - X_i\|_F^2. \quad (3)$$

In general, no closed-form solution exists for this problem, so we proposed an iterative algorithm in [22] and [23] to acquire a local optimum. The main difference between [22] and [23] is that the former is only for matrix representations while the latter is for general tensors of arbitrary order. The solution to the above formulation is obtained as follows. First, for a given $U \in \mathbb{R}^{m \times m'}$, the objective function (3) can be rewritten as

$$V^* = \arg\min_V \sum_i \left\|X_i^U VV' - X_i\right\|_F^2 \quad (4)$$

where $X_i^U = UU'X_i$. As demonstrated in [22], the solution of (4) is the first $n'$ leading eigenvectors of the eigenvalue decomposition problem $FF'x = \lambda x$ with

$$F = \left[X_1^U(1,*)', \ldots, X_1^U(m,*)',\ X_2^U(1,*)', \ldots, X_N^U(m,*)'\right] \quad (5)$$

where $X_i^U(r,*)$ is the $r$th row vector of the image matrix $X_i^U$ and $F$ is the concatenated matrix of all $X_i^U$.

Similarly, for a given $V \in \mathbb{R}^{n \times n'}$, the optimization problem (3) becomes

$$U^* = \arg\min_U \sum_i \left\|UU'X_i^V - X_i\right\|_F^2 \quad (6)$$

where $X_i^V = X_iVV'$. As demonstrated in [22], the solution of (6) is the first $m'$ leading eigenvectors of the eigenvalue decomposition problem $GG'x = \lambda x$ with

$$G = \left[X_1^V(*,1), \ldots, X_1^V(*,n), X_2^V(*,1), \ldots, X_N^V(*,n)\right] \quad (7)$$

where $X_i^V(*,c)$ is the $c$th column vector of image matrix $X_i^V$. By iteratively optimizing the objective function with respect to $U$ and $V$, respectively, we can obtain a local optimum. More details on the above deduction and its optimization can be found in [22] and [23].

*2) DATER:* DATER [24] is a supervised learning algorithm directly based on a general tensor representation. Its objective function with a matrix representation is

$$(U^*, V^*) = \arg\max_{U,V} \frac{\sum_{c=1}^{N_c} n_c \|U'\overline{X}_cV - U'\overline{X}V\|_F^2}{\sum_{i=1}^{N} \left\|U'X_iV - U'\overline{X}_{l_i}V\right\|_F^2} \quad (8)$$

where $\overline{X}_c$ is the average matrix of the samples belonging to the $c$th class, $\overline{X}$ is the average matrix for all the samples, and $n_c$ is the sample number of the $c$th class. In (8), the between-class scatter (numerator) is measured by the sum of the weighted distances between the dimensionality reduced class centers and the total sample center, while the within-class scatter (denominator) is measured by the sum of the distances between the dimensionality reduced samples and their corresponding class average matrices.

Again, there is no closed-form solution for (8). Thus, in [24], we proposed an iterative scheme. For a given $U \in \mathbb{R}^{m \times m'}$, the objective function of (8) is rewritten as

$$\frac{\sum_{c=1}^{N_c} n_c \|U'\overline{X}_cV - U'\overline{X}V\|_F^2}{\sum_{i=1}^{N} \left\|U'X_iV - U'\overline{X}_{l_i}V\right\|_F^2} = \frac{\sum_{c=1}^{N_c} n_c \left\|\overline{X}_c^uV - \overline{X}^uV\right\|_F^2}{\sum_{i=1}^{N} \left\|X_i^uV - \overline{X}_{l_i}^uV\right\|_F^2} = \frac{\mathrm{Tr}\left(V'S_b^uV\right)}{\mathrm{Tr}\left(V'S_w^uV\right)} \quad (9)$$

where $\mathrm{Tr}(\mathrm{A})$ denotes the trace of a matrix A, $X_i^u = U'X_i$, $\overline{X}_c^u = U'\overline{X}_c$, $\overline{X}^u = U'\overline{X}$, and

$$S_b^u = \sum_c n_c \left(\overline{X}_c^u - \overline{X}^u\right)' \left(\overline{X}_c^u - \overline{X}^u\right)$$
$$S_w^u = \sum_i \left(X_i^u - \overline{X}_{l_i}^u\right)' \left(X_i^u - \overline{X}_{l_i}^u\right). \quad (10)$$

The matrix $V$ with given $U$ is obtained by the generalized eigenvalue decomposition [1] method by transforming the objective function to $\mathrm{Tr}((V'S_w^uV)^{-1}(V'S_b^uV))$ as

$$S_b^uv_i = \lambda_i S_w^uv_i \quad (11)$$

where $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n'-1}$ are the $n'$ largest eigenvalues; $v_i$ is the eigenvector corresponding to eigenvalue $\lambda_i$, which constitutes the $i$th column of the solution $V$.

Similarly, for a given $V \in \mathbb{R}^{n\times n'}$, the objective function of (8) is changed to

$$\frac{\sum_{c=1}^{N_c} n_c \|U'\overline{X}_c V - U'\overline{X}V\|_F^2}{\sum_{i=1}^{N} \left\|U'X_i V - U'\overline{X}_{l_i}V\right\|_F^2} = \frac{\sum_{c=1}^{N_c} n_c \left\|U'\overline{X}_c^v - U'\overline{X}^v\right\|_F^2}{\sum_{i=1}^{N} \left\|U'X_i^v - U'\overline{X}_{l_i}^v\right\|_F^2} = \frac{\mathrm{Tr}\left(U'S_b^v U\right)}{\mathrm{Tr}\left(U'S_w^v U\right)} \tag{12}$$

where $X_i^v = X_i V$, $\overline{X}_c^v = \overline{X}_c V$, $\overline{X}^v = \overline{X}V$, and

$$S_b^v = \sum_c n_c \left(\overline{X}_c^v - \overline{X}^v\right)\left(\overline{X}_c^v - \overline{X}^v\right)'$$
$$S_w^v = \sum_i \left(X_i^v - \overline{X}_{l_i}^v\right)\left(X_i^v - \overline{X}_{l_i}^v\right)'. \tag{13}$$

Again, the matrix $U$ with a given $V$ is calculated using the generalized eigenvalue decomposition method [1] by transforming the objective function to $\mathrm{Tr}((U'S_w^v U)^{-1}(U'S_b^v U))$ as

$$S_b^v u_i = \lambda_i S_w^v u_i \tag{14}$$

where $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{m'-1}$ are the $m'$ largest eigenvalues; $u_i$ is the eigenvector corresponding to eigenvalue $\lambda_i$, which constitutes the $i$th column of the solution $U$.

*3) Analysis of Learnability and Computation Complexity of Matrix-Based Representation:* We take CSA as an example to illustrate the learnability of matrix-based representations compared with the vector-based algorithm PCA. In PCA, the feature dimension of the object is $mn$ and the total number of samples is $N$. For example, the dimensionality of average silhouettes in the USF HumanID database is above 10 000 ($128 * 88$) and the available number of average silhouette images for model training is fewer than 1000. Hence, it is quite difficult for PCA to robustly acquire the real representative components. However, for CSA, as shown in (5), each row vector is considered as an object to be analyzed. Hence, the dimension of the row vector is the width $n$ of the image $X_i^U$, and the total number of objects is $Nm$. Also, in (7), each column is used as the object for eigenvalue decomposition, so the dimension of the object is the height $m$ and the total number of objects is $Nn$. With more training data and a lower feature dimension, CSA can better discover the representative information, and the typical small sample size problem and the curse of dimensionality are alleviated, which results in CSA's superiority over traditional PCA.

The computational complexity of CSA is lower than that of PCA when the total number of samples $N$ is comparable in magnitude to the feature dimension $mn$. For ease of understanding, we assume $m = n$. Then, the complexity of PCA is $\mathrm{O}(\mathrm{n}^6)$,[1] while CSA's complexity is $\mathrm{O}(\mathrm{n}^3)$ for each substep. Although CSA iteratively obtains the solution, it is still faster than PCA owing to its simplicity in each substep optimization.

[1]Although eigenvalue decomposition can be conducted on a covariance matrix of size $N \times N$, when $N$ is much less than $mn$, we can consider the complexity of PCA as $\mathrm{O}(\mathrm{n}^6)$ in the case that $N$ is comparable in magnitude to $mn$.

### B. Gait Recognition With Matrix Representation

Here, we apply the previous matrix-representation-based algorithms for gait recognition. In this study, we focus on the learning algorithm for gait recognition applications, and thus we start our analysis from the binary image sequences. As shown in [6] and [11], the complete sequence is partitioned into several subsequences according to the gait period length $N_{\mathrm{Gait}}$, which is provided by Sarkar *et al.* [13] in the USF HumanID database. For each sequence, the binary silhouette images within a single gait cycle are averaged to acquire several gray-level average silhouette images $AS_i$ by

$$AS_i = \frac{1}{N_{\mathrm{Gait}}} \sum_{k=(i-1)N_{\mathrm{Gait}}+1}^{k=iN_{\mathrm{Gait}}} S(k),\ i = 1, \ldots, \lfloor T/N_{\mathrm{Gait}} \rfloor \tag{15}$$

where $S = \{S(1), \ldots, S(T)\}$ denotes the binary images for one sequence, with $T$ as the total number of frames; $\lfloor T/N_{\mathrm{Gait}} \rfloor$ denotes the floor function, which is the largest integer that is not larger than $T/N_{\mathrm{Gait}}$.

In the USF HumanID database, the training set $\{X_i \in \mathbb{R}^{m\times n}, i = 1, \ldots, N\}$ comprise all of the gray-level average silhouette images from all of the sequences in the gallery set, and the test set comprises the average silhouette images from all of the sequences in the related probe set. We note that, in the USF HumanID database, there are no overlapped sequences between the Gallery and Probe sets.

We utilize average silhouettes for gait recognition because of the following observations. First, they are more robust than the time-series-based silhouette representation [13], especially since binary images from the current object segmentation algorithms are quite noisy. Second, the average silhouette images do not depend on the choice of the starting stance of the gait cycle, and thus gait alignment [13] is unnecessary. Fig. 1 shows some original binary images and the average silhouettes of the first person in the Gallery set and Probe Set A. The average silhouettes are gray-level images, and darker regions are more likely to be the pedestrian. From it, we can observe that a person may have similar average silhouette images in varying instances.

Since the average silhouette images are in the form of matrices, our previous algorithms CSA and DATER can be used as the learning algorithms for model training. Inspired by the classical two-stage PCA+LDA scheme [1], [6], we combine CSA and DATER, namely CSA+DATER, as the learning algorithm. As in PCA+LDA, in CSA+DATER, CSA is applied to remove noise and retain the most representative information, and DATER is used for extracting information to distinguish different persons. The only difference between CSA+DATER with PCA+LDA is that CSA and DATER are both based on matrix representations. The detailed algorithm is listed in Fig. 2.

In the testing stage, with $U$ and $V$, each gray-level average silhouette image in the gallery and probe sets are converted into

Given the training sample set $\{X_i \in \mathbb{R}^{m\times n},\ i=1,...,N\}$, denote the iteration numbers $T_c^{max}$ and $T_d^{max}$ in CSA and DATER, the lower dimensions $(m_c', n_c')$ in the CSA step and $(m_d', n_d')$ in the DATER step. Note that for dimensionality reduction, we have $m_d' \le m_c'$ and $n_d' \le n_c'$.

1. Initialization: set $U_c^0 = I_m$ or arbitrary columnly orthogonal matrix.

2. For t=1, 2,…, $T_c^{max}$, Do

   a) For a given $U_c^{t-1}$, compute the optimal projection matrix $V_c^t$ according to (4) and (5).

   b) For a given $V_c^t$, compute the optimal projection matrix $U_c^t$ according to (6) and (7).

   c) If t > 1 and $\|U_c^t - U_c^{t-1}\|_F < m * \varepsilon_c$ and $\|V_c^t - V_c^{t-1}\|_F < n * \varepsilon_c$, go to step 3.

3. The projection matrices in the CSA step are $U_c = U_c^t \in \mathbb{R}^{m\times m_c'}$ and $V_c = V_c^t \in \mathbb{R}^{n\times n_c'}$.

4. Compute the projected samples $\{Y_i \in \mathbb{R}^{m_c'\times n_c'},\ i=1,...,N\}$, $Y_i = U_c' X_i V_c$ and use them as the new training data in step 6.

5. Initialization: $U_d^0 = I_{m_c'}$ or an arbitrary columnly orthogonal matrix.

6. For t=1, 2,…, $T_d^{max}$, Do

   a) For a given $U_d^{t-1}$, compute the optimal projection matrix $V_d^t$ according to (11).

   b) For a given $V_d^t$, compute the optimal projection matrix $U_d^t$ according to (14).

   c) If t > 1 and $\|U_d^t - U_d^{t-1}\|_F < m_c' * \varepsilon_d$ and $\|V_d^t - V_d^{t-1}\|_F < n_c' * \varepsilon_d$, go to step 7.

7. The projection matrices in the DATER step is $U_d = U_d^t \in \mathbb{R}^{m_c'\times m_d'}$ and $V_d = V_d^t \in \mathbb{R}^{n_c'\times n_d'}$.

8. Output the final projection matrices as $U = U_c U_d$ and $V = V_c V_d$.

Fig. 2. Training procedure of CSA+DATER.

a lower dimensional matrix. Considering that the median operation is more robust to noise effects than the traditional minimum operation, we use the same distance measure for gallery sequences and probe sequences as in [11], [13]

$$\mathrm{Dist}(LS_P, LS_G) = \mathrm{Median}_{i=1}^{N_P}\left(\min_{j=1}^{N_G} \|LS_P(i) - LS_G(j)\|_F\right) \tag{16}$$

where $\{LS_P(i), i = 1,\ldots,N_P\}$ and $\{LS_G(j), j = 1,\ldots,N_G\}$ are the lower dimensional average silhouette image matrices of one gallery sequence and one probe sequence, respectively, with $N_P$ and $N_G$ as the total number of average silhouette images, and $\mathrm{Dist}(LS_P, LS_G)$ is the distance between the two sequences. Based on the distance between the two sequences, the nearest neighbor classifier is used to make the final decision.

## IV. Experiments

Our experiments are carried out on the USF HumanID gait database [13]. For real gait data, we list the results of the baseline algorithm [13] and Han and Bhanu's work [6] in Table II,

TABLE II
RANK-1 AND RANK-5 RECOGNITION ACCURACY (%) OF BASELINE ALGORITHM [13], HAN AND BHANU'S WORK [6], DATER [24], AND CSA+DATER ON THE USF HUMANID GAIT DATABASE

| | Rank-1 Performance | | | | Rank-5 Performance | | | |
|---|---|---|---|---|---|---|---|---|
| Set | Ref [13] | Ref [6] | DATER | CSA+DATER | Ref [13] | Ref [6] | DATER | CSA+DATER |
| A | 73 | 87 | 87 | 89 | 88 | 92 | 96 | 96 |
| B | 78 | 85 | 93 | 93 | 93 | 93 | 96 | 96 |
| C | 48 | 76 | 78 | 80 | 78 | 89 | 93 | 94 |
| D | 32 | 31 | 42 | 44 | 66 | 58 | 69 | 74 |
| E | 22 | 30 | 42 | 45 | 55 | 60 | 69 | 79 |
| F | 17 | 18 | 23 | 25 | 42 | 36 | 51 | 53 |
| G | 17 | 21 | 28 | 33 | 38 | 43 | 52 | 57 |
| H | 61 | 63 | 80 | 80 | 85 | 90 | 92 | 93 |
| I | 57 | 59 | 79 | 79 | 78 | 81 | 90 | 91 |
| J | 36 | 54 | 59 | 60 | 62 | 79 | 83 | 83 |
| K | 3 | 3 | 18 | 18 | 12 | 12 | 40 | 40 |
| L | 3 | 6 | 21 | 21 | 15 | 12 | 36 | 36 |

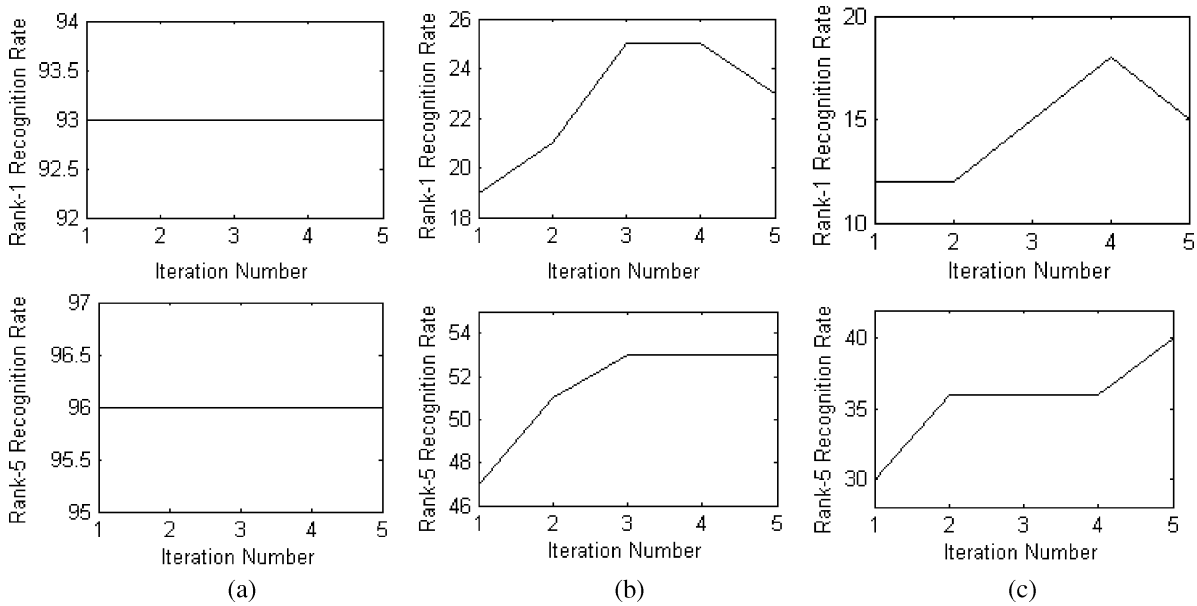


Fig. 3. Illustration of the Rank-1 and Rank-5 performance variations with different iteration number $T_d^{\max}$. Note that for computational simplicity, we set $T_c^{\max} = T_d^{\max}$. From this, it is easy to see that recognition rates change slightly for Probe Sets B, F, and K if the iteration number is greater than 3. (a) Rank-1 and Rank–5 performance of B. (b) Rank-1 and Rank–5 performance of F. (c) Rank-1 and Rank–5 performance of K.

where Rank-1 indicates that the correct subject is ranked as the top candidate and Rank-5 means that the correct subject is ranked among the top five candidates. We also report the results of DATER [24] for comparison. Note that, for a fair comparison, we use the same binary silhouettes and gait period length for all algorithms.

It can be observed that DATER [24] outperforms the state-of-the-art algorithms [6], [13] in most cases, which highlights the effectiveness of our matrix representation for gait recognition. We also observe that generally, the CSA+DATER algorithm outperforms DATER and achieves the best performance, which demonstrates the contribution of CSA as the preprocessing step before DATER.

In our study, like in several other gait recognitions algorithms [6], [19], [20], several parameters need to be decided. In our previous research [23], [24], we have observed that $T_c^{\max}$ and $T_d^{\max}$ have little effect on the final recognition performance in both CSA and DATER, so we consider only one parameter $T_d^{\max}$ of the two and set $T_c^{\max} = T_d^{\max}$. We observe that $T_d^{\max}$ affects the final recognition rate slightly if the iteration number is greater than 3. We illustrate this for Probe Sets B, F, and K in Fig. 3. Considering that the highest recognition rates are acquired with larger iteration numbers in some probe sets, we have run the experiments with $T_d^{\max}$ ranging from 1 to 5 in this work and report

[2][Online]. Available: http://marathon.csee.usf.edu/GaitBaseline

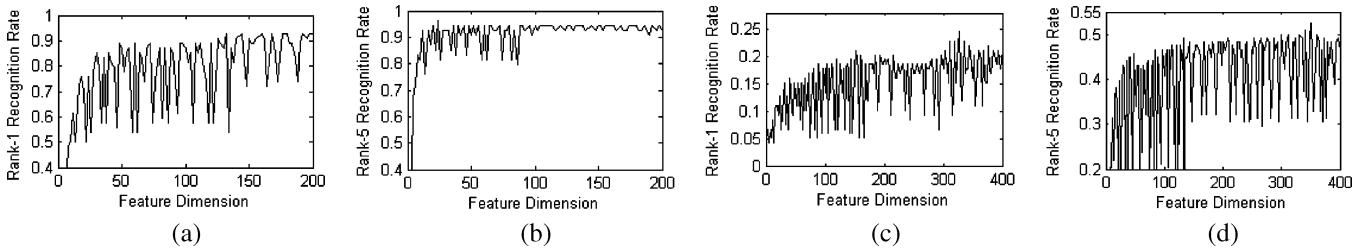


Fig. 4. Illustration of the Rank-1 and Rank-5 performance variation with different configuration $(m'_d, n'_d)$. The horizontal axis is the product of $m'_d$ and $n'_d$. In this figure, we set the parameters $(T_c^{\max}, T_d^{\max})$ as (3, 3), and $(m'_c, n'_c)$ as (70, 70), (45, 45), (100, 88), and (70, 70) for each figure. respectively. (a) Rank-1 performance of B. (b) Rank-5 performance of B. (c) Rank-1 performance of F. (d) Rank-5 performance of F.

the best results. However, as mentioned, we observe that even with a larger fixed iteration number (e.g., 3), the performance is slightly degraded. Considering the computational burden, we set the lower dimension parameter configuration as follows.

1) By setting $n'_c = m'_c$, we only need to consider one parameter $m'_c$. In this work, $m'_c$ varies by intervals of five and a starting dimension of 20. For the USF HumanID database, the width $n$ (88) is less than the height $m$ (128), so, for the case that $m'_c$ is greater than $n$, we fix $n'_c$ as $n$. Also, in this study, we consider one special configuration with $m'_c = m$ and $n'_c = n$, i.e., we preserve 100% of the energy in CSA stage. In this case, CSA+DATER is equal to DATER, so CSA+DATER cannot be worse than DATER in theory
2) With $m'_c$, we run all possible configurations of $(m'_d, n'_d)$.

In Fig. 4, we plot the Rank-1 and Rank-5 performance variation with different configuration $(m'_d, n'_d)$ for Probe Sets B and K. It is unclear how to choose the optimal parameter configuration for new test data in theory, and we still do not know how to choose the optimal $m'_d, n'_d$ even with the fixed $(m'_c, n'_c)$. How to determine these parameters is a potential direction for future work.

## V. Conclusion

We have applied the matrix-representation-based subspace learning algorithms on a biometric problem that is of growing research interest—human gait recognition. Each sequence was represented as several gray-level average silhouette images, which correspond to gait cycles. A two-stage scheme called CSA+DATER has been employed for dimensionality reduction directly based on the matrix representations. The experiments on the standard USF HumanID Gait database demonstrate encouraging performance improvements over the state-of-the-art algorithms [6], [13] for human gait recognition.

## Acknowledgment

The authors would like to thank Dr. S. Lin for proofreading this paper.

## References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[2] C. BenAbdelkader and R. Cutler, "Motion-based recognition of people in eigengait space," in *Proc. IEEE Conf. Automat. Face Gesture Recogn.*, 2002, pp. 254–259.

[3] A. Bobick and A. Johnson, "Gait recognition using static activity-specific parameters," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2001, pp. 423–430.

[4] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, 2002, pp. 351–356.

[5] D. Cunado, M. Nixon, and J. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Computer Vision and Image Understanding*, vol. 90, pp. 1–41, Apr. 2003.

[6] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 842–847.

[7] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. Chowdhury, V. Kruger, and R. Chellappa, "Identification of humans using gait," *IEEE Trans. Image Process.*, vol. 13, pp. 1163–1173, Sep. 2004.

[8] L. Lee and W. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, 2002, pp. 148–155.

[9] L. Lee, G. Dalley, and K. Tieu, "Learning pedestrian models for silhouette refinement," in *Proc. IEEE Conf. Computer Vision*, 2003, pp. 663–670.

[10] Z. Liu, L. Malave, and S. Sarkar, "Studies on silhouette quality and gait recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 704–711.

[11] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: averaged silhouette," in *Proc. IEEE Conf. Pattern Recognit.*, 2004, pp. 211–214.

[12] S. Mowdray and M. Nixon, "Extraction and recognition of periodically deforming objects by continuous, spatio-temporal shape description," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 895–901.

[13] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The HumanID gait challenge problem: data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.

[14] R. Tanawongsuwan and A. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2001, pp. 726–731.

[15] ——, "Modelling the effects of walking speed on appearance-based gait recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 783–789.

[16] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 1991, pp. 586–591.

[17] A. Veeraraghavan, A. Chowdhury, and R. Chellappa, "Role of shape and kinematics in human movement analysis," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 730–737.

[18] G. Veres, L. Gordon, J. Carter, and M. Nixon, "What image information is important in silhouette-based gait recognition?," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2004, pp. 776–782.

[19] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1120–1131, Sep. 2003.

[20] ——, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.

[21] ——, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Feb. 2004.

[22] D. Xu, S. Yan, L. Zhang, Z. Liu, and H. Zhang, Coupled Subspaces Analysis Microsoft Res. Tech. Rep. MSR-TR-2004-106 [Online]. Available: http://research.microsoft.com/research/pubs/view.aspx?type=Technical%20Report&id=811

[23] D. Xu, S. Yan, L. Zhang, H. Zhang, Z. Liu, and H. Shum, "Concurrent subspace analysis," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2005, pp. 203–208.

[24] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang , and H. Zhang, "Discriminant analysis with tensor representation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2005, pp. 526–532.

[25] J. Ye, "Generalized low rank approximations of matrices," in *Proc. Int. Conf. Machine Learning*, 2004.

[26] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems 17*. Cambridge, MA, MIT Press, 2005, pp. 1569–1576.