Abstract—This paper describes a method of discriminant analysis for cross-view recognition with a relatively small number of training samples. Since appearance of a recognition target (e.g., face, gait, gesture, and action) is in general drastically changes as an observation view changes, we introduce multiple view-specific projection matrices and consider to project a recognition target from a certain view by a corresponding view-specific projection matrix into a common discriminant subspace. Moreover, conventional vectorized representation of an originally higher-order tensor object (e.g., a spatio-temporal image in gait recognition) often suffers from the curse of dimensionality dilemma, we therefore encapsulate the multiple view-specific projection matrices in a framework of discriminant analysis with tensor representation, which enables us to overcome the curse of dimensionality dilemma. Experiments of cross-view gait recognition with two publicly available gait databases show the effectiveness of the proposed method in case where a training sample size is small.

I. INTRODUCTION

Recognition from different views, namely, cross-view recognition, has been one of central topics in the computer vision and pattern recognition communities for a long time, since view changes are naturally observed in many applications (e.g., face, gait, gesture, and action recognition) and also induce drastic appearance changes of a target.

For this purpose, view-specific projections to a common subspace, are considered in many studies to cope with large appearance changes by view changes. The most popular way to obtain a common subspace for multiple views is canonical correlation analysis (CCA) [7], [6], which learns two projection matrices for a set of two variables so as to maximize a correlation between them in the common subspace. In additions, several variants of CCA have been also proposed, such as kernel CCA (KCCA) [2] and sparse CCA [5]. Whereas the above approach only consider to analyze pairwise variables, multi-view CCA (MCCA) [20] was proposed to obtain one common space for more than two views, where multiple view-specific transforms were obtained by maximizing the total correlation between any pairs of views.

Moreover, a family of regression is also regarded as one of view-specific approaches. Partial least squares (PLS) regression [24], [19] projects samples from two views to a common latent subspace, where samples from one view are regarded as regressor while those from the other view as regressand. For example, PLS is employed for face recognition with pose, low-resolution, and sketch in [21]. Support vector regression (SVR) [23] is an extension from support vector machine (SVM) to regression problem and it is employed in cross-view gait recognition [11] for example.

Although all the above methods could maximize correlation (or minimize differences) among two or more views, they do not take discrimination aspect into consideration. A straightforward solution is to employ discriminant analysis. A typical example is linear discriminant analysis (LDA) [18] which project an object with a single view-common matrix into a lower dimensional discriminant subspace in a supervised way, where a between-class variance is maximized and a within-class variance is minimized at the same time. It is, however, difficult in essence to efficiently mitigate the intra-class variance induced by view changes with a single view-common matrix.

On the other hand, discriminative approaches with multiple view-specific projections have been proposed. As extensions from CCA, correlation discriminant analysis (CDA) [15] and discriminative CCA (DCCA) [10] are proposed, where within-class correlation is maximized while between-class correlation is minimized. Moreover, as extensions from LDA, multi-view fisher discriminant analysis (MFDA) [3] for binary classification problem, and generalized multi-view analysis (GMLDA) [22] for multi-class classification from multiple views are proposed. While GMLDA requires hyper-parameter setting for regularization, multi-view discriminant analysis (MvDA) [9] provide more direct derivation from LDA for multiple view-specific projection matrices without any hyper parameters. In addition, MvDA simultaneous obtains a concatenation of multiple view-specific projection matrices by solving a single generalized eigenvalue problem in an analytical way.

On the other hand, despite that an object handled in computer vision and pattern recognition often has an higher-order tensor structure originally such as an image (a second-order tensor, namely, a matrix) and a voxel volume or spatio-temporal image (a third-order tensor), most of the above subspace learning approaches first vectorize the original tensor object into the first-order tensor, namely, a vector without keeping the original structure and thereafter project it into a lower-dimensional subspace. Such a first-order tensor of feature vector usually has a considerably high dimensionality (e.g., an image with 640 by 480 pixel size leads to a

This work was partly supported by JSPS Grants-in-Aid for Scientific Research (S) 21220003, Young Scientists (A) 23680017, “R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society”, Strategic Funds for the Promotion of Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government, and the JST CREST “Behavior Understanding based on Intention-Gait Model” project.
vector of 307,200 dimensions). Such a considerably higher dimensional vector often induces the curse of dimensionality dilemma or small sample size problem through subspace learning stage. More specifically, a within-class scatter matrix used in many discriminant analysis approaches, is in general singular (degenerated) in particular in case where the size of training samples is small.

In order to overcome the problem, Yan et al. [28] propose discriminant analysis with tensor representation (DATER) which treats an original tensor object as is rather than vectorizing it into a first-order tensor with high dimensionality. In DATER, multiple projection matrices are prepared for each mode, more specifically, \( L \) projection matrices for an \( L \)-order tensor object, and such mode-specific projection matrices are optimized in turn. Since the dimension considered in each optimization is at most the number of components in each mode (e.g., 640 and 480 for an image with 640 by 480 pixel size for the first and second modes, respectively) while the number of training samples is multiplied by the number of components in the other modes, DATER significantly mitigates the curse of dimensionality dilemma or small sample size problem. As an example of DATER application, Xu et al. [26] applies it to gait energy image (GEI) [4], which is a second-order tensor object, and show the effectiveness of DATER in human gait recognition problem. DATER may, however, still suffer from large within-class variations by view changes since it uses a single view-common projection matrix for each mode.

We therefore propose a method of multi-view discriminant analysis with tensor representation (MvDATER) by considering both advantages of enhanced discrimination capability by view-specific projections and tolerance to the small sample size problem by mode-specific projections. More specifically, we prepare multiple mode-specific and view-specific projection matrices (i.e., \( L_{NV} \) projection matrices for an \( L \)-order tensor from \( N_V \) views). We then encapsulate MvDA algorithm [9] in DATER algorithm [28] where multiple view-specific projection matrices for a certain mode is optimized through so-called \( l \)-mode discriminant analysis and where those for another mode is optimized in turn. Note that MvDATER is not a trivial combination of two existing approaches, i.e., MvDATER is beyond sequential application of MvDA and DATER, since it is a unified formulation obtained by extending the state-of-the-art MvDA into the tensor domain in a technically sound way.

Compared with previous works, the proposed method simultaneously satisfy the following properties: (1) A single common discriminative subspace is obtained for multiple views by jointly optimizing multiple view-specific projection matrices (see Fig. 1). (2) Optimization for each mode discriminant analysis is solved analytically through generalized eigenvalue problem. Since individual generalized eigenvalue problem is solved with much smaller dimensions, (3) computational cost for each mode discriminant analysis is reduced to a large extent and also (4) the curse of dimensionality dilemma is avoided. (5) The small sample size problem is overcome since the sample size is effectively multiplied by a large scale as described later.

II. MULTI-VIEW DISCRIMINANT ANALYSIS WITH TENSOR REPRESENTATION

A. Tensor representation

Target objects in computer vision are often represented as a second or higher order tensor rather than a vector. For example, a single image and an image sequence (video) are represented as a second order tensor (matrix) and a third-order tensor, respectively. Most of the conventional approaches to subspace learning, such as principal component analysis (PCA), LDA, CCA, MvDA, firstly unfold the tensor object into a vector object (e.g., an image object \( X \in \mathbb{R}^{H \times W} \) with the height \( H \) and the width \( W \) is unfolded into a vector \( x \in \mathbb{R}^{HW} \) with the image size dimensionality \( HW \), and then
derive lower dimensional subspace from the vector object. As a result, they often suffer from the curse of dimension and/or the small sample size problem. We therefore derive a lower dimensional subspace directly from the tensor objects so as to keep the original data structure as well as avoid the curse of dimension and the small sample size problem.

More specifically, we consider an $L$-th order tensor object $A \in \mathbb{R}^{M_{1} \times \ldots \times M_{L}}$, where $l$-mode dimensionality is represented by $M_{l}$ and whose component is represented using $L$ indices $\{m_{l}\}$ as $A(m_{1}, \ldots, m_{L})$. Note that the total number of components in the tensor object $A$ sums up to $M = \prod_{l=1}^{L} M_{l}$.

In the following, we further review a couple of basic techniques of tensor algebra [13].

**Inner product, norm, and distance:** The inner product (or scalar product) of two tensors $A, B \in \mathbb{R}^{M_{1} \times \ldots \times M_{L}}$ with the same dimensionality is defined as

$$
(A, B) = \sum_{m_{1}=1}^{M_{1}} \ldots \sum_{m_{L}=1}^{M_{L}} A(m_{1}, \ldots, m_{L})B(m_{1}, \ldots, m_{L}).
$$

The Frobenius norm of a tensor $A$ is defined as $\|A\|_{F} = \sqrt{(A, A)}$, and subsequently a distance between two tensors $A, B$ is defined as $d(A, B) = \|A - B\|_{F}$.

**The $l$-mode product:** The $l$-mode product of a tensor $A$ by a matrix $U_{l} \in \mathbb{R}^{M_{l} \times M'_{l}}$, denoted as $A \times_{l} U_{l}$, is a tensor $B \in \mathbb{R}^{M_{1} \times \ldots \times M_{l-1} \times M'_{l} \times M_{l+1} \times \ldots \times M_{L}}$ whose component is

$$
B(m_{1}, \ldots, m_{l-1}, m'_{l}, m_{l+1}, \ldots, m_{L}) = \sum_{m_{l}=1}^{M_{l}} A(m_{1}, \ldots, m_{L})U_{l}(m_{l}, m'_{l}).
$$

Note that the $l$-mode product of tensor changes the $l$-mode dimensionality of the tensor from $M_{l}$ to $M'_{l}$ while keeping the dimensionality of the other modes.

**The $l$-mode vectors and unfolding:** The $l$-mode vectors of a tensor $A$ are defined as a set of $M_{l}$-dimensional vectors obtained from the tensor $A$ by varying its index $m_{l}$ while keeping the other indices $\{m_{p}\} (p \neq l)$ fixed as illustrated in Fig. 2. Since the total number of the $l$-mode vectors sums up to $T^{(l)} = \prod_{p \neq l} M_{p}$, the $l$-mode unfolding of the tensor $A$ is represented a matrix $A^{(l)} = [a^{(l)}, 1, \ldots, a^{(l)}, T^{(l)}] \in \mathbb{R}^{M_{l} \times T^{(l)}}$ whose column vectors $\{a^{(l)}, t\} (t = 1, \ldots, T^{(l)})$ are the $l$-mode vectors. In this paper, we refer to the $l$-mode unfolding operation as $A^{(l)} \leftarrow_{l} A$ and note that a bracketed superscript $(l)$ indicates notation for the $l$-mode unfolding in this paper for the convenience. We also note that the norm of the $l$-mode product is rewritten using the $l$-mode unfolding by considering simple algebra as

$$
\|A \times_{l} U_{l}\|_{F} = \left\| \sum_{t=1}^{T^{(l)}} (a^{(l)}, t) \left( U_{l} \right)_{tj} \right\|_{F} = \left\| (A^{(l)})^{T} U_{l} \right\|_{F}.
$$

**B. Multi-view projections**

Since the objective of the discriminant analysis is to find lower-dimensional discriminant subspace, we consider a multi-linear projection from an original tensor $X \in \mathbb{R}^{M_{1} \times \ldots \times M_{L}}$ into a lower-dimensional but the same-order tensor $Y \in \mathbb{R}^{M'_{1} \times \ldots \times M'_{L}} (M'_{l} < M_{l} \forall l)$ as

$$
Y = X \times_{1} U_{1} \ldots \times_{L} U_{L},
$$

where $U_{l} \in \mathbb{R}^{M_{l} \times M'_{l}}$ is a projection matrix for the $l$-mode product.

Although conventional approaches to discriminant analysis such as LDA consider a common projections regardless of the difference of data domains, it is in general difficult to find efficient common projections in case where tensor objects as features (e.g., face images or gait image sequences) are significantly different among the domains (e.g., different views in face or gait recognition).

We therefore introduce a multi-domain multi-mode projections to overcome such domain differences at the same time to keep tensor object structures, unlike the DATER [28] considers only the multi-mode aspect and the MvDA [9] does only the multi-domain aspect. Although we refer to the domain as view after this in accordance with the MvDA framework [9], note that the proposed framework is applicable to not only the view domain but also a variety of domains (e.g., illumination and expression in face recognition, walking speed and clothing in gait recognition).

Formally, we define the multi-view multi-mode projection matrices as $U = \{U_{l,j} \in \mathbb{R}^{M_{l} \times M'_{l}}\} (l = 1, \ldots, L, j = 1, \ldots, N_{V})$, where subscripts $l$ and $j$ indicate mode and view, respectively, and $N_{V}$ is the number of views. We then project a tensor object from any view into a common discriminant subspace by switching the projection matrix based on the domain where the tensor object $X$ comes from accordingly, as shown in Fig. 1.

**C. Discriminant tensor criterion**

In this subsection, we introduce a discriminant tensor criterion with the multi-view multi-mode projection matrices. For this purpose, formally, let us define a set of training tensor objects $X = \{X_{i,j,k} \in \mathbb{R}^{M_{1} \times \ldots \times M_{L}}\} (i = 1, \ldots, N_{C}, j = 1, \ldots, N_{V}, k = 1, \ldots, n_{ij})$, where $X_{i,j,k}$ is the $k$-th training tensor object of the $i$-th class from the $j$-th view, and $N_{C}$ and $n_{ij}$ are the number of classes and the number of training samples of the $i$-th class from the $j$-th view, respectively. We subsequently define the number of training tensor objects of the $i$-th class as $n_{i} = \sum_{j=1}^{N_{V}} n_{ij}$ and also the total number of training tensor objects as $n = \sum_{i=1}^{N_{C}} n_{i}$.

Since the training tensor object $X_{i,j,k}$ comes from the $j$-th view, the corresponding lower-dimensional tensor object

![Fig. 2. Illustration of the 1-mode vectors and unfolding.](image-url)
where 

$$Y_{ijk} \in \mathbb{R}^{M_1 \times \ldots \times M_L}$$

is the common discriminant subspace represented as

$$Y_{ijk} = X_{ijk} \times_1 U_{1,j} \ldots \times_L U_{L,j}. \quad (5)$$

Here, a set of multi-view multi-mode projection matrices are optimized by maximizing a between-class scatter while minimizing a within-class scatter, namely, by maximizing their ratio as

$$U^* = \arg \max_U \frac{\sum_{i=1}^{NC} n_i \| Y_i - \bar{Y} \|^2_F}{\sum_{i=1}^{NC} \sum_{j=1}^{M_1} \sum_{k=1}^{M_L} \| Y_{ijk} - \bar{Y}_i \|^2_F}, \quad (6)$$

where the denominator and the numerator are the within-class and between-class scatter, respectively, and 

$$\bar{Y}_i \in \mathbb{R}^{M_1 \times \ldots \times M_L}$$

and 

$$\bar{Y} \in \mathbb{R}^{M_1 \times \ldots \times M_L}$$

are the i-th class mean and the total mean, respectively. The i-th class mean 

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad (7)$$

The total mean $\bar{Y}$ is similarly derived as

$$\bar{Y} = \sum_{i=1}^{NC} w_i \bar{Y}_i = \sum_{i=1}^{NC} \sum_{j=1}^{n_i} w_{ij} \bar{X}_{ij} \times_1 U_{1,j} \ldots \times_L U_{L,j}, \quad (9)$$

Now, by substituting Eqs. (5)(7)(9) into Eq. (6), we obtain

$$U^* = \arg \max_U \frac{\sum_{i=1}^{NC} n_i \sum_{j=1}^{n_i} w_{ij} \| \bar{X}_{ij} \times_1 U_{1,j} \ldots \times_L U_{L,j} \|_F^2}{\sum_{i=1}^{NC} \sum_{j=1}^{n_i} \sum_{k=1}^{M_L} \| \bar{X}_{ijk} \times_1 U_{1,j} \ldots \times_L U_{L,j} \|_F^2}. \quad (10)$$

Since there is in general no closed-form solution for Eq. (10) due to the higher-order tensor structure, we alternatively search for an iterative solution to derive the common discriminant subspace as described in subsection II-E.

D. l-mode discriminant analysis

In this subsection, we introduce an l-mode discriminant analysis, which is an essential technique for the iterative solution described in subsection II-E. More specifically, we consider another discriminant criterion focused only on the l-mode product with projection matrices $U_l = \{ U_{l,j} \in \mathbb{R}^{M_1 \times M_L} \} (j = 1, \ldots, N_l)$ as

$$U_{l,*} = \arg \max_{U_l} \frac{\text{Tr} \left( \sum_{r=1}^{NC} \sum_{i=1}^{N} \sum_{j=1}^{N_r} U_{l,j}^T S_{B,r}^{(l)} U_{l,j} \right) T}{\sum_{r=1}^{NC} \sum_{i=1}^{N} \sum_{j=1}^{N_r} \| X_{ij} \times_1 U_{1,j} \ldots \times_L U_{L,j} \|_F^2}. \quad (11)$$

Recall that the norm of the l-mode product is represented by the l-mode unfolding as Eq. (3), we reformulate Eq. (11) as (please refer to supplementary material for the detailed derivation)

$$U_{l,*} = \arg \max_{U_l} \frac{\text{Tr} \left( \sum_{r=1}^{NC} \sum_{i=1}^{N} \sum_{j=1}^{N_r} U_{l,j}^T S_{B,r}^{(l)} U_{l,j} \right) T}{\sum_{r=1}^{NC} \sum_{i=1}^{N} \sum_{j=1}^{N_r} \| X_{ij} \times_1 U_{1,j} \ldots \times_L U_{L,j} \|_F^2}. \quad (12)$$

where $\text{Tr}(\cdot)$ means a trace of a matrix, and $S_{W,r}^{(l)} \in \mathbb{R}^{M_1 \times M_l}$ and $S_{W,r}^{(l)} \in \mathbb{R}^{M_1 \times M_l}$ are within-class and between-class scatter matrices, respectively, for l-mode unfolding from a pair of the r-th view and s-th view which are defined as

$$S_{W,r}^{(l)} = \sum_{i=1}^{N} \sum_{j=1}^{N_r} \sum_{k=1}^{M_l} \| X_{ij} \times_1 U_{1,j} \ldots \times_L U_{L,j} \|_F^2$$

where $\delta_{rs}$ is Kronecker’s delta, and $X_{ij}^{(l)} \in \mathbb{R}^{M_l \times T(l)}$ and $X_{ij}^{(l)} \in \mathbb{R}^{M_l \times T(l)}$ are the l-mode unfolding of $X_{ij}$ and $\bar{X}_{ij}$, i.e., $X_{ij}^{(l)} \leftarrow X_{ij}$ and $\bar{X}_{ij}^{(l)} \leftarrow \bar{X}_{ij}$, respectively. Moreover, we can rewrite Eq. (12) by introducing view concatenated version of matrices as

$$U_{l,*} = \arg \max_{U_l} \frac{\text{Tr} (U_{l}^T S_{B}^{(l)} U_{l})}{\text{Tr} (U_{l}^T S_{W}^{(l)} U_{l})}, \quad (15)$$

where $U_l \in \mathbb{R}^{N_l \times M_1 \times M_l}$ is the l-mode view-concatenated projection matrix, and $S_{W}^{(l)} \in \mathbb{R}^{N_l \times M_1 \times M_l}$ and $S_{B}^{(l)} \in \mathbb{R}^{N_l \times M_1 \times M_l}$ are view-concatenated within-class and between-class scatter matrices for the l-mode unfolding, which are respectively defined as

$$S_{W}^{(l)} = \left[ S_{W,11}^{(l)} \ldots S_{W,1N_l}^{(l)} \right], \quad S_{B}^{(l)} = \left[ S_{B,11}^{(l)} \ldots S_{B,1N_l}^{(l)} \right], \quad \text{Tr} (U_{l}^T S_{B}^{(l)} U_{l})$$

(16)
Since the closed form solution for the objective function in Eq. (17), which is in the form of trace ratio, does not exist according to [25], we relax the objective function into a more tractable one in the form of ratio trace as bellow:

\[ U_{i*} = \arg \max_{U_i} \text{Tr} \left( U_i^T S_B^{(l)} U_i \right) / \text{Tr} \left( U_i^T S_W^{(l)} U_i \right), \]

which can be solved analytically through generalized eigenvalue problem:

\[ S_B^{(l)} U_i = S_W^{(l)} U_i \Lambda, \]

where \( \Lambda \in \mathbb{R}^{M_L \times M_L} \) is an orthogonal matrix whose diagonal components are eigenvalues. We then extract the first \( M_L \) largest eigenvectors as a solution to \( U_{i*} \). For more detailed discussion on the number of available projection directions, we refer the reader to the literature [28] due to page limitation.

E. Iterative solution

As described before, since the discriminant tensor criterion defined by Eq. (10) often has no closed-form solution, we introduce an iterative solution to this. More specifically, we defined by Eq. (10) often has no closed-form solution, we refer the reader to the literature [28] due to page limitation.

F. Computational complexity

In this subsection, we discuss the properties of the proposed method in terms of computational complexity compared with closely related approaches such as LDA [18], MvDA [9], DATER [28] to the proposed method, MvDATER. For simplicity of the analysis, we consider a situation where each mode has the same dimensionality (i.e., \( M_l = M \forall l \)), and where the number of the training samples per class per view is the same (i.e., \( n_{ij} = n_T \forall i,j \)). In addition, since the generalized eigenvalue problem is the most important part w.r.t. the computational complexity, we focus on the generalized eigenvalue problem.

LDA: The \( L \)-order tensor is vectorized when computing the scatter matrices, and the dimensionality of the scatter matrix is then \( \prod_{l=1}^{L} M_l^2 \). Since the dimensionality \( M_l^2 \) is considerably high, it is often the case that the number of training samples is much less than the dimensionality \( n_l \ll M_l^2 \), which results in singularity of the within-class scatter matrix (small sample size problem, curse of the dimensionality). Moreover, the computational complexity for the generalized eigenvalue problem is cubic order of the dimensionality. LDA for the higher-order tensor objects requires high computational cost \( O \left((M^3)^3 \right) \).

MvDA: Since MvDA constructs a view-concatenated scatter matrix composed of \( N_v \times N_v \) sub-matrices (see Eq. (16) for reference), the dimensionality is \( N_v \)-time larger than that for LDA (i.e., \( N_v M^3 \)). In addition, the number of training samples needs to be considered at the submatrix level in essence, and it is hence \( N_v \)-time smaller than that for LDA (i.e., \( n_v / N_v \)). Although MvDA has a good discrimination capability for multiple views, it more suffers from the singularity problem than LDA and also higher computational cost \( O \left((N_v M^3)^3 \right) \).

DATER: Since DATER constructs a scatter matrix through unfolding operation for each mode, the mode-wise dimensionality is just \( M_l \), which is much smaller than that of LDA \( M^L \). In addition, since the unfolding operation also drastically increase the number of training samples as \( N_T \prod_{l \neq l} M_l = N M^L-1 \). Therefore, DATER mitigates the small sample size problem to large extent because the condition, \( N M^L-1 > M \), holds in most case. Moreover, the computational cost for each mode and loop is \( O(M^3) \) and hence that for the whole process is at most \( O(N_{iter} L M^3) \).

MvDATER: In analogous to relation between LDA and MvDA, the dimensionality is \( N_v \)-time larger than that for DATER (i.e., \( N_v M \)), while the number of training samples is \( N_v \)-time smaller than that for DATER (i.e., \( n_v / N_v \)). As a result, the computational cost is \( (N_v M^3)^3 \)-time larger than that for DATER (i.e., \( O \left((N_{iter} L (N_v M))^3 \right) \)). The number of views \( N_v \) is, however, much smaller than the date dimension \( M \) in general (e.g., \( N_v = 2 \) when handling pairwise view), we can say that the proposed MvDATER realize a reasonable tradeoff among discrimination capability, small sample size problem, and the computational cost.

III. APPLICATION TO CROSS-VIEW GAIT RECOGNITION

A. Setup

We evaluated the proposed MvDATER approach under cross-view gait recognition (i.e. gait-based person authentication) using the most prevailing gait feature, i.e., GEI [4], with two publicly available gait databases: (1) CASIA Gait Database B (call it CASIA later) [29] and (2) the OU-ISIR Gait Database, the Large Population Data set (call it OU-LP later) [8]. CASIA contains walking sequences from
Algorithm 1 MvDATER

Input: A set of $L$-order training tensor objects $X = \{X_{ijk} \in \mathbb{R}^{M_1 \times \ldots \times M_L}\}$ ($i = 1, \ldots, N_C, j = 1, \ldots, N_V, k = 1, \ldots, n_{ij}$), a set of dimensions in the common discriminant subspace $\{M'_l\}$, convergence criteria $\epsilon$, and the maximum iteration $N_{iter}$

Output: A set of projection matrices $U = \{U_{l,j} \in \mathbb{R}^{M_1 \times M'_l}\}$ ($l = 1, \ldots, L, j = 1, \ldots, N_V$)

1: $U^{prev}_{l,j}, U^{cur}_{l,j} \leftarrow ISM \forall l, j$ \hspace{1cm} $\triangleright$ Initialization
2: for $iter = 1$ to $N_{iter}$ do
3: \hspace{1cm} for $l = 1$ to $L$ do
4: \hspace{2cm} $Y_{ijk} \leftarrow X_{ijk} \times U^{cur}_{l-1,j} \times \ldots \times U^{cur}_{l-1,j+1} U^{prev}_{l,j+1} \times \ldots \times U^{prev}_{L,j}$ \hspace{1cm} $\triangleright$ The $l$-mode unfolding
5: \hspace{2cm} $Y_{ijk} \leftarrow \frac{1}{n} \sum_{i,j} y_{ijk} \forall i, j$
6: \hspace{2cm} $S^{(l)}_{W,rs} \leftarrow \sum_{i=1}^{N_C} \left\{ \sum_{k=1}^{n_i} \delta_{is} (Y^{(l)}_{ik})^T \left( \frac{\mu_{i}^{(l)}}{\sigma_{i}^{(l)}} \right) \left( Y^{(l)}_{i} \right)^T \right\} \forall r, s$
7: \hspace{2cm} $S^{(l)}_{W,is} \leftarrow \sum_{i=1}^{N_C} \frac{n_{is}}{n} \\sum_{k=1}^{n_i} \delta_{is} (Y^{(l)}_{ik})^T \left( Y_{is}^{(l)} \right)^T \forall s$
8: \hspace{2cm} $U^{cur}_{l,j} \leftarrow \left( \frac{1}{n_i} \sum_{k=1}^{n_i} \sum_{i=1}^{N_C} (Y_{i}^{(l)T}) (Y_{is}^{(l)T}) \right) \left( \sum_{i=1}^{N_C} (Y_{i}^{(l)T}) \right) \forall i, j$
9: \hspace{1cm} $\triangleright$ Within-class and between-class scatter matrices
10: \hspace{2cm} Update $U^{cur}_{l,j}$ by solving $S^{(l)}_{W} U_{l,j} = S^{(l)}_{W} U_{l,j}$ \hspace{1cm} $\triangleright$ Generalized eigenvalue problem
11: \hspace{2cm} end for
12: \hspace{1cm} if $||U^{cur}_{l,j} - U^{prev}_{l,j}||_F < M_l M'_l \epsilon \forall l$ then
13: \hspace{2cm} \hspace{1cm} break \hspace{1cm} $\triangleright$ Convergence condition
14: \hspace{2cm} end if
15: \hspace{2cm} $U^{prev}_{l,j} \leftarrow U^{cur}_{l,j}$ \hspace{1cm} $\triangleright$ Update
16: \hspace{1cm} end for
17: Output $\{U^{cur}_{l,j}\}$

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensionality of the Scatter Matrices, the Number of Training Samples, and the Computational Complexity W.R.T. the Generalized Eigenvalue Problem</td>
</tr>
<tr>
<td>Dimensionality</td>
</tr>
<tr>
<td>$#$ Training samples</td>
</tr>
<tr>
<td>Complexity</td>
</tr>
</tbody>
</table>

124 subjects captured from wide range of views (i.e., 18° intervals from 0° (frontal view) to 180° (rear view), and hence it is suitable to evaluate gait recognition under large view variations. We divided the subjects into disjoint sets of 62 test subjects and 62 training subjects. For the test subjects, we set the first normal walking sequences from a view 90° as a gallery, while we set the second sequence from views 55°, 65°, 75°, 85° as probes, respectively. Similar to CASIA, training was done for pairwise view. Examples of the GEIs (44 by 64 pixel-size) at six different views can be seen in Fig. 3.

As for performance measures, we picked up rank-1 identification rates (denoted as Rank-1 later) a.k.a. correct classification rate (CCR) in identification scenarios (i.e. one-to-many matching) as well as equal error rate (EER) of false acceptance rate of imposters (different persons) and false rejection rate of genuine (the same person) in verification scenarios (i.e. one-to-one matching). We compared the proposed MvDATER with three closely related approaches as benchmarks: LDA [18], MvDA [9], and DATER [28], and dissimilarity features are computed by Euclidean distance in each discriminant space. Note that each benchmark is followed by the preprocessing dimension reduction approaches, more specifically, PCA for LDA and MvDA, and concurrent subspace analysis (CSA) [27] for DATER and MvDATER.
### Results for CASIA (#training subjects: 62)

The proposed MvDATER achieved the best or the second best accuracies in many cases.

<table>
<thead>
<tr>
<th>Probe view</th>
<th>Rank-1</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72°</td>
<td>54°</td>
</tr>
<tr>
<td>LDA</td>
<td>80.3%</td>
<td>29.4%</td>
</tr>
<tr>
<td>MvDA</td>
<td>4.2%</td>
<td>4.2%</td>
</tr>
<tr>
<td>DATER</td>
<td>56.5%</td>
<td>9.0%</td>
</tr>
<tr>
<td>MvDATER</td>
<td>66.5%</td>
<td>48.1%</td>
</tr>
</tbody>
</table>

### Results for OU-LP (#training subjects: 10)

As an exception, trained projection matrices for LDA and MvDA did not perform well in low-dimensional discriminant subspaces. While the DATER overcome such a troublesome small sample size problem, it still suffers from insufficient discrimination capability because it only has a single view-common projection for each mode, which results in poor performance for large view variations.

<table>
<thead>
<tr>
<th>Probe view</th>
<th>Rank-1</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75°</td>
<td>65°</td>
</tr>
<tr>
<td>LDA</td>
<td>12.0%</td>
<td>5.1%</td>
</tr>
<tr>
<td>MvDA</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>DATER</td>
<td>61.3%</td>
<td>25.7%</td>
</tr>
<tr>
<td>MvDATER</td>
<td>67.2%</td>
<td>12.6%</td>
</tr>
</tbody>
</table>

#### CASIA

To show the robustness of the proposed method against small sample size problem, we picked up only one normal sequence per subject from the training set and trained the proposed MvDATER as well as the other benchmarks. As shown in Table II, we can see that the proposed MvDATER achieved the best or the second best for almost all the settings. Because within-class scatter matrices for LDA and MvDA suffers from singularity (in particular for MvDA) due to small sample size problem, trained projection matrices for LDA and MvDA did not perform well in low-dimensional discriminant subspaces. While the DATER overcome such a troublesome small sample size problem, it still suffers from insufficient discrimination capability because it only has a single view-common projection for each mode, which results in poor performance for large view variations. On the other hand, the proposed MvDATER has multiple view-specific projection matrices for each mode and at the same time avoids the small sample size problem, and it therefore outperforms the other benchmarks as a result. As an exception, LDA outperforms MvDATER for 72° probe. This is because gait features from 90° view and 72° view are relatively similar each other and hence even a single projection matrix can successfully absorb the intra-class variations among them.

#### OU-LP

Whereas the samples per subject was limited in the previous experiment, we limited the number of training subjects to 10 in this experiment to check the robustness against small sample size problem. As shown in Table III, we can see that the proposed MvDATER performs well on average, although it does not work well for view 55°. In addition, DATER seems to be comparable to MvDATER. This is because the OU-LP contains much larger variation in test subjects than that of CASIA and hence DATER, which is the most robust to small training sample sizes, performs relatively well.

In order to further investigate the effect of the number of training subjects, we show the performance transition against the number of training subjects for view 75° from OU-LP in Fig. 5. From this graph, we observe the followings. (1) LDA and MvDA perform well for sufficient number of training subjects (e.g., more than 100 subjects), and their performance drastically drop as the number of training subjects decreases. In particular, MvDA, that is the most recent benchmark, performs quite poorly when training sample sizes are small. This reveals the limitation of MvDA and prompt us to more focus on the generalization capability aspect in future avenue of research. (2) DATER keeps its performance against the decrease of the number of training subjects, although its basic performance is lower than the other benchmarks. From another perspective, we can say that it does not increase the performance as the training sample sizes increase. (3) MvDATER exhibit higher performance than DATER thanks to multiple view-specific projections and keeps relatively good performance against the decrease of the number of training subjects. As a result, we can confirm the strength of the proposed method in case where the number of training samples is small.

#### IV. Discussion

##### A. Connections to 2DLDA and its variant

As also discussed in the literature [28], while 2DLDA [12] only considers a 2-mode discriminant analysis, DATER considers discriminant analysis for all the modes in turn. More specifically, the 2DLDA is formulated as a special case of MvDATER with Nv = 1, L = 2, and U_{11} = I. In addition, a straightforward multi-view extension of the 2DLDA, that is, 2DMvDA, could be considered, and it is again formulated as...
a special case of MvDATER with \( NV = 2, L = 2, \) and \( U_{1,j} = I \) \((j = 1, 2)\). The proposed MvDATER is therefore regarded as a unified framework for these discriminant analyses.

**B. Class masking problem**

Since the proposed MvDATER is built upon the LDA which optimizes the Bayes error for the case of unimodal Gaussian classes with equal covariances, it might increase the overlap between the class conditional densities in the lower dimensional subspace in a heteroscedastics setting, which is so-called class masking problem. To cope with the class masking problem, Moustafa et al. [1] employed pareto discriminant analysis which simultaneously maximizes each class-pairwise distance and which thus encourages the case that all classes are equidistant from each other in the lower dimensional space. Since the pareto discriminant analysis can be encapsulate in each \( l \)-mode discriminant analysis of the proposed MvDATER, we will extend the MvDATER so as to mitigate the class masking problem in future.

**V. CONCLUSION**

This paper described a method of multi-view discriminant analysis with tensor representation (MvDATER) for cross-view recognition with a relatively small number of training samples. We introduce multiple view-specific and mode-specific projection matrices so as that high-order tensor objects from multiple views can be projected into a single common discriminant subspace. In the proposed algorithm, multiple view-specific projection matrices are jointly and analytically optimized via a single generalized eigenvalue problem with smaller dimension for each mode, which draws many of the advantages such as efficient cross-view handling and overcomng the curse of dimensionality dilemma and small sample size problem.

While we validated the effectiveness of the proposed method with cross-view gait recognition problems compared with the most relevant benchmarks, we will further compare it with more advanced approaches to cross-view gait recognition (e.g., [17], [16], [11], [14]). Moreover, we will further validate it with a variety of cross-view recognition such as action recognition and face recognition in future.

**REFERENCES**


