

A Survey of Multilinear Subspace Learning for Tensor Data

Haiping Lu^{a*}, K. N. Plataniotis^b, A. N. Venetsanopoulos^{b,c}

^a*Institute for Infocomm Research, Agency for Science, Technology and Research,
#21-01 Connexis (South Tower), 1 Fusionopolis Way, Singapore 138632*

^b*The Edward S. Rogers Sr. Department of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, ON, M5A 3G4, Canada*

^c*Department of Electrical and Computer Engineering, Ryerson University, 350
Victoria Street, Toronto, ON, M5B 2K3, Canada*

Abstract

Increasingly large amount of multidimensional data are being generated on a daily basis in many applications. This leads to a strong demand for learning algorithms to extract useful information from these massive data. This paper surveys the field of multilinear subspace learning (MSL) for dimensionality reduction of multidimensional data directly from their tensorial representations. It discusses the central issues of MSL, including establishing the foundations of the field via multilinear projections, formulating a unifying MSL framework for systematic treatment of the problem, examining the algorithmic aspects of typical MSL solutions, and categorizing both unsupervised and supervised MSL algorithms into taxonomies. Lastly, the paper summarizes a wide range of MSL applications and concludes with perspectives on future research directions.

Key words: Subspace learning, dimensionality reduction, feature extraction, multidimensional data, tensor, multilinear, survey, taxonomy

Please cite:

Haiping Lu, K. N. Plataniotis and A. N. Venetsanopoulos, "A Survey of Multilinear Subspace Learning for Tensor Data", *Pattern Recognition*, vol. 44, no. 7, pp. 1540-1551, Jul. 2011.

BibTeX:

<http://www.dsp.utoronto.ca/~haiping/BibTeX/MSLSurvey2011.bib>

* Corresponding author. Tel: (65)64082527; Fax: (65)67761378 Email: hplu@ieee.org

Article published in Pattern Recognition 44 (2011) 1540–1551

1 Introduction

With the advances in data collection and storage capabilities, massive multidimensional data are being generated on a daily basis in a wide range of emerging applications, and learning algorithms for knowledge extraction from these data are becoming more and more important. Two-dimensional (2D) data include gray-level images in computer vision and pattern recognition [1–4], multichannel EEG signals in biomedical engineering [5, 6], and gene expression data in bioinformatics [7]. Three-dimensional (3D) data include 3D objects in generic object recognition [8], hyperspectral cube in remote sensing [9], and gray-level video sequences in activity or gesture recognition for surveillance or human-computer interaction (HCI) [10, 11]. There are also many multidimensional signals in medical image analysis [12], content-based retrieval [1, 13], and space-time super-resolution [14] for digital cameras with limited spatial and temporal resolution. In addition, many streaming data and mining data are frequently organized as third-order tensors [15–17]. Data in environmental sensor monitoring are often organized in three modes of time, location, and type [17]. Data in social network analysis are usually organized in three modes of time, author, and keywords [17]. Data in network forensics are often organized in three modes of time, source, and destination, and data in web graph mining are commonly organized in three modes of source, destination, and text [15].

These massive multidimensional data are usually very high-dimensional, with a large amount of redundancy, and only occupying a subspace of the input space [18]. Thus, for feature extraction, dimensionality reduction is frequently employed to map high-dimensional data to a low-dimensional space while retaining as much information as possible [18, 19]. However, this is a challenging problem due to the large variability and complex pattern distribution of the input data, and the limited number of samples available for training in practice [20]. Linear subspace learning (LSL) algorithms are traditional dimensionality reduction techniques that represent input data as vectors and solve for an optimal linear mapping to a lower dimensional space. Unfortunately, they often become inadequate when dealing with massive multidimensional data. They result in very high-dimensional vectors, lead to the estimation of a large number of parameters, and also break the natural structure and correlation in the original data [2, 21, 22].

Due to the challenges in emerging applications above, there has been a pressing need for more effective dimensionality reduction schemes for massive multidimensional data. Recently, interests have grown in multilinear subspace learning (MSL) [2, 21–26], a novel approach to dimensionality reduction of multidimensional data where the input data are represented in their natural multidimensional form as tensors. Figure 1 shows two examples of tensor data

representations for a face image and a silhouette sequence. MSL has the potential to learn more compact and useful representations than LSL [21,27] and it is expected to have potential future impact in both developing new MSL algorithms and solving problems in applications involving massive multidimensional data. The research on MSL has gradually progressed from heuristic exploration to systematic investigation [28] and both unsupervised and supervised MSL algorithms have been proposed in the past a few years [2,21–26].

It should be noted that MSL belongs to tensor data analysis (or tensor-based computation and modeling), which is more general and has a much wider scope. Multilinear algebra, the basis of tensor data analysis, has been studied in mathematics for several decades [29–31] and there are a number of recent survey papers summarizing recent developments in tensor data analysis. E.g., Qi *et al.* review numerical multilinear algebra and its applications in [32]. Muti and Bourennane [33] survey new filtering methods for multicomponent data modelled as tensors in noise reduction for color images and multicomponent seismic data. Acar and Yener [34] surveys unsupervised multiway data analysis for discovering structures in higher-order data sets in applications such as chemistry, neuroscience, and social network analysis. Kolda and Bader [35] provide an overview of higher-order tensor decompositions and their applications in psychometrics, chemometrics, signal processing, etc. These survey papers primarily focus on unsupervised tensor data analysis through factor decomposition. In addition, Zafeiriou [36] provides an overview of both unsupervised and supervised nonnegative tensor factorization (NTF) [37,38] with NTF algorithms and their applications in visual representation and recognition discussed.

In contrast, this survey paper focuses on a systematic introduction to the field of MSL. To the best knowledge of the authors, this is the first unifying survey of both unsupervised and supervised MSL algorithms. For detailed introduction and review on multilinear algebra, multilinear decomposition, and NTF, the

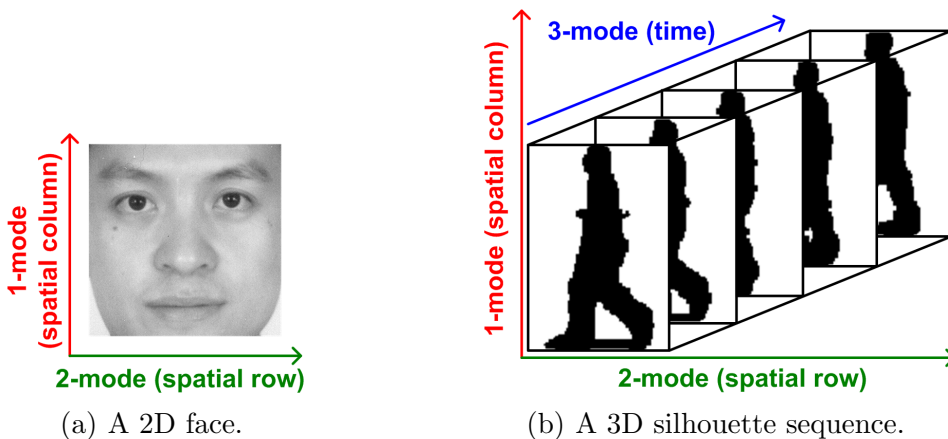


Fig. 1. Illustration of real-world data in their natural tensor representation.

Table 1
List of Important Acronyms.

Acronym	Description
EMP	Elementary Multilinear Projection
LDA	Linear Discriminant Analysis
LSL	Linear Subspace Learning
MSL	Multilinear Subspace Learning
PCA	Principal Component Analysis
SSS	Small Sample Size
TTP	Tensor-to-Tensor Projection
TVP	Tensor-to-Vector Projection
VVP	Vector-to-Vector Projection

readers should go to [30–36, 39, 40] while this paper serves as a complement to these surveys. In the rest of this paper, Section 2 first introduces notations and basic multilinear algebra, and then addresses multilinear projections for direct mapping from high-dimensional tensorial representations to low-dimensional vectorial or tensorial representations. This section also provides insights to the relationships among different projections. Section 3 formulates a unifying MSL framework for systematic treatment of the MSL problem. A typical approach to solve MSL is presented and several algorithmic issues are examined in this section. Under the MSL framework, existing MSL algorithms are reviewed, analyzed and categorized into taxonomies in Section 4. Finally, MSL applications are summarized in Section 5 and future research topics are covered in Section 6. For easy reference, Table 1 lists several important acronyms used in this paper.

2 Fundamentals and Multilinear Projections

This section first reviews the notations and some basic multilinear operations [30, 31, 41] that are necessary in defining the MSL problem. The important concepts of multilinear projections are then introduced, including elementary multilinear projection (EMP), tensor-to-vector projection (TVP), and tensor-to-tensor projection (TTP), and their relationships are explored. Table 2 summarizes the important symbols used in this paper for quick reference.

Table 2
List of Important Notations

Notation	Description
I_n	the (input) dimensionality of the n -mode
M	the number of training samples
N	the order of a tensor object, the number of indices/modes
P	the number of EMPs in a TVP
P_n	the n -mode dimensionality in the projected space of a TTP
$\mathbf{U}^{(n)}$	the n th projection matrix
$\mathbf{u}_p^{(n)^T}$	the n -mode projection of the p th EMP
$\text{vec}(\cdot)$	the vectorized representation
\mathcal{X}	an input tensor sample
\mathcal{Y}	the projection of \mathcal{X} on $\{\mathbf{U}^{(n)}\}$
$\mathbf{y}(p)$	the projection of \mathcal{X} on $\{\mathbf{u}_p^{(n)^T}, n = 1, \dots, N\}$
$\ \cdot\ _F$	the Frobenius norm

2.1 Notations

This paper follows the notation conventions in multilinear algebra, pattern recognition, and adaptive learning literature [30, 31, 41]. Vectors are denoted by lowercase boldface letters, e.g., \mathbf{x} ; matrices by uppercase boldface, e.g., \mathbf{X} ; and tensors by calligraphic letters, e.g., \mathcal{X} . Their elements are denoted with indices in parentheses. Indices are denoted by lowercase letters, spanning the range from 1 to the uppercase letter of the index, e.g., $p = 1, 2, \dots, P$. In addressing part of a vector/matrix/tensor, “:” denotes the full range of the respective index and $n_1 : n_2$ denotes indices ranging from n_1 to n_2 . In this paper, only real-valued data are considered.

2.2 Basic Multilinear Algebra

As in [30, 31, 41], an N th-order tensor is denoted as: $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, which is addressed by N indices i_n , $n = 1, \dots, N$, with each i_n addressing the n -mode of \mathcal{A} . The n -mode product of a tensor \mathcal{A} by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted as $\mathcal{A} \times_n \mathbf{U}$, is a tensor with entries [30]:

$$(\mathcal{A} \times_n \mathbf{U})(i_1, \dots, i_{n-1}, j_n, i_{n+1}, \dots, i_N) = \sum_{i_n} \mathcal{A}(i_1, \dots, i_N) \cdot \mathbf{U}(j_n, i_n). \quad (1)$$

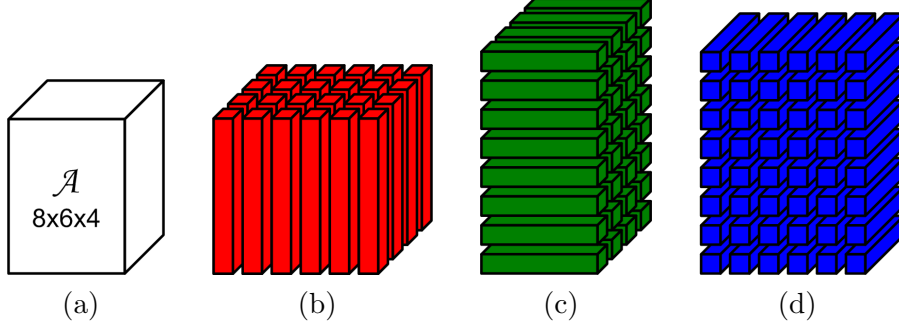


Fig. 2. Visual illustration of n -mode vectors: (a) a tensor $\mathcal{A} \in \mathbb{R}^{8 \times 6 \times 4}$, (b) the 1-mode vectors, (c) the 2-mode vectors, and (d) the 3-mode vectors.

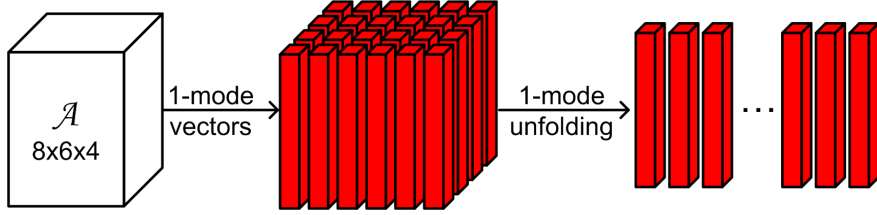


Fig. 3. Visual illustration of 1-mode unfolding of a third-order tensor.

The scalar product of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is defined as:

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} \mathcal{A}(i_1, i_2, \dots, i_N) \cdot \mathcal{B}(i_1, i_2, \dots, i_N) \quad (2)$$

and the Frobenius norm of \mathcal{A} is defined as [30]

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}. \quad (3)$$

The “ n -mode vectors” of \mathcal{A} are defined as the I_n -dimensional vectors obtained from \mathcal{A} by varying its index i_n while keeping all the other indices fixed. A rank-one tensor \mathcal{A} equals to the outer product of N vectors:

$$\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)}, \quad (4)$$

i.e.,

$$\mathcal{A}(i_1, i_2, \dots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdot \dots \cdot \mathbf{u}^{(N)}(i_N) \quad (5)$$

for all values of indices. Unfolding \mathcal{A} along the n -mode is denoted as

$$\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}, \quad (6)$$

where the column vectors of $\mathbf{A}_{(n)}$ are the n -mode vectors of \mathcal{A} . Figures 2(b), 2(c), and 2(d) illustrate the 1-mode, 2-mode, and 3-mode vectors of a tensor \mathcal{A} in Fig. 2(a), respectively. Figure 3 shows the 1-mode unfolding of the tensor \mathcal{A} in Fig. 2(a).

The distance between tensors \mathcal{A} and \mathcal{B} can be measured by the Frobenius norm [2]:

$$dist(\mathcal{A}, \mathcal{B}) = \| \mathcal{A} - \mathcal{B} \|_F . \quad (7)$$

Although this is a tensor-based measure, it is equivalent to a distance measure of corresponding vector representations, as proven in [42]. Let $vec(\mathcal{A})$ be the vector representation (vectorization) of \mathcal{A} , then

$$dist(\mathcal{A}, \mathcal{B}) = \| vec(\mathcal{A}) - vec(\mathcal{B}) \|_2 . \quad (8)$$

This implies that the distance between two tensors as defined in (7) equals to the Euclidean distance between their vectorized representations.

2.3 Multilinear Projections

A multilinear subspace is defined through a multilinear projection that maps the input tensor data from one space to another (lower-dimensional) space [43]. Therefore, multilinear projection is a key concept in MSL. There are three basic multilinear projections based on the input and output of a projection: the traditional vector-to-vector projection (VVP), TTP, and TVP.

2.3.1 Vector-to-Vector Projection

Linear projection is a standard transformation used widely in various applications [44, 45]. A linear projection takes a vector $\mathbf{x} \in \mathbb{R}^I$ and projects it to another vector $\mathbf{y} \in \mathbb{R}^P$ using a projection matrix $\mathbf{U} \in \mathbb{R}^{I \times P}$:

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} = \mathbf{x} \times_1 \mathbf{U}^T . \quad (9)$$

In typical pattern recognition applications, $P \ll I$. Therefore, linear projection is a VVP. When the input to VVP is an N th-order tensor \mathcal{X} with $N > 1$, it needs to be reshaped into a vector as $\mathbf{x} = vec(\mathcal{X})$ before projection. Figure 4(a) illustrates the VVP of a tensor object \mathcal{A} . Besides the traditional linear projection, there are alternative ways to project a tensor to a low-dimensional space, as shown in Fig. 4(b), which will be discussed below.

2.3.2 Tensor-to-Tensor Projection

A tensor can be projected to another tensor of the same order, named as TTP. As an N th-order tensor \mathcal{X} resides in the tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ [30, 43], the tensor space can be viewed as the Kronecker product of N vector (linear) spaces $\mathbb{R}^{I_1}, \mathbb{R}^{I_2}, \dots, \mathbb{R}^{I_N}$ [43]. To project a tensor \mathcal{X} in a tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ to another tensor \mathcal{Y} in a lower-dimensional tensor space

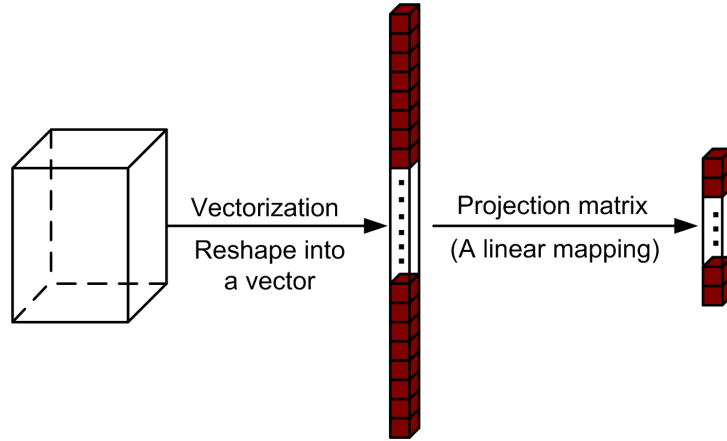
$\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \dots \otimes \mathbb{R}^{P_N}$, where $P_n \leq I_n$ for all n , N projection matrices $\{\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, \dots, N\}$ are used so that [31]

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \dots \times_N \mathbf{U}^{(N)T}. \quad (10)$$

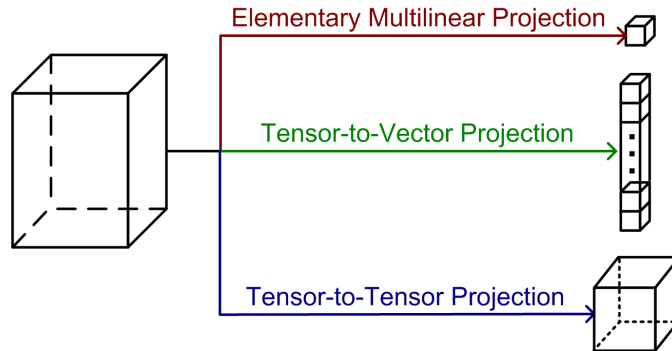
It can be done in N steps, where in the n th step, each n -mode vector is projected to a lower dimension P_n by $\mathbf{U}^{(n)}$, as shown in Fig. 5(a). Figure 5(b) demonstrates how to project a tensor in 1-mode using a 1-mode projection matrix, which projects each 1-mode vector of the original tensor to a low-dimensional vector.

2.3.3 Tensor-to-Vector Projection

The third multilinear projection is TVP, which is referred to as the rank-one projections in some works [46–48]. It projects a tensor to a vector, which can be viewed as multiple projections from a tensor to a scalar, as illustrated in Fig. 6(a), where the TVP of a tensor $\mathcal{A} \in \mathbb{R}^{8 \times 6 \times 4}$ to a $P \times 1$ vector consists of P projections from \mathcal{A} to a scalar. Thus, the projection from a tensor to a



(a) Vector-to-vector (linear) projection.



(b) Alternative ways to project a tensor.

Fig. 4. Ways to project a tensor to a low-dimensional space.

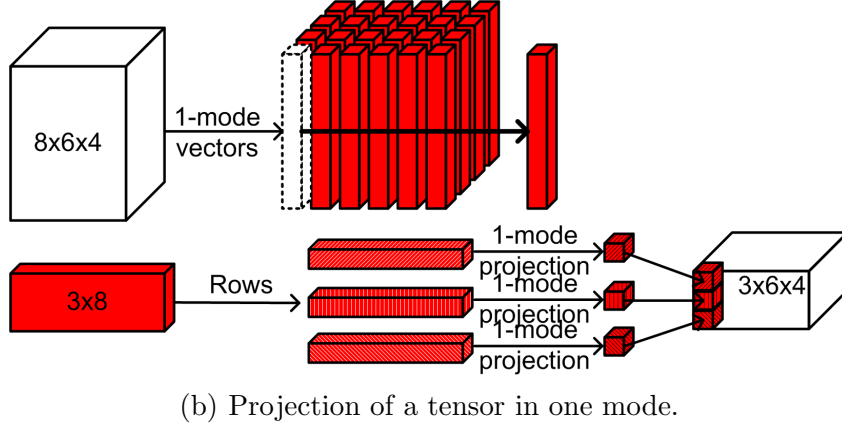
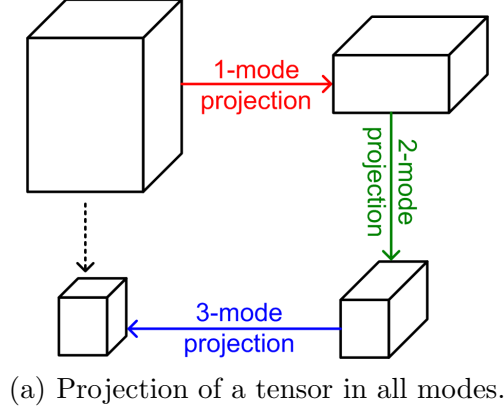


Fig. 5. Illustration of tensor-to-tensor projection.

scalar is considered first.

A tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be projected to a point y through N unit projection vectors $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ as:

$$y = \mathcal{X} \times_1 \mathbf{u}^{(1)T} \times_2 \mathbf{u}^{(2)T} \dots \times_N \mathbf{u}^{(N)T}, \quad \|\mathbf{u}^{(n)}\| = 1 \text{ for } n = 1, \dots, N, \quad (11)$$

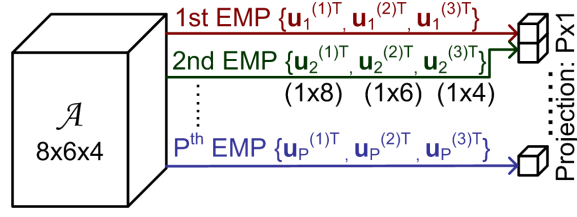
where $\|\cdot\|$ is the Euclidean norm for vectors. It can be written in the scalar product (2) as:

$$y = \langle \mathcal{X}, \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)} \rangle. \quad (12)$$

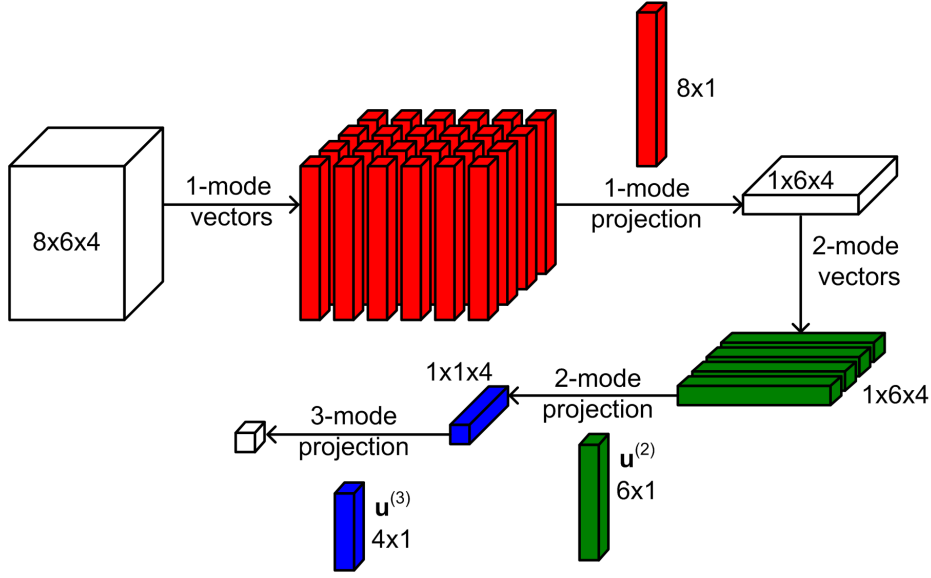
Denote $\mathcal{U} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)}$, then $y = \langle \mathcal{X}, \mathcal{U} \rangle$. This multilinear projection $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ is named as an EMP, the projection of a tensor on a single line (resulting a scalar), with one projection vector in each mode. Figure 6(b) illustrates an EMP of a tensor.

Thus, the TVP of a tensor object \mathcal{X} to a vector $\mathbf{y} \in \mathbb{R}^P$ in a P -dimensional vector space consists of P EMPs $\{\mathbf{u}_p^{(1)T}, \mathbf{u}_p^{(2)T}, \dots, \mathbf{u}_p^{(N)T}\}, p = 1, \dots, P$, which can be written concisely as $\{\mathbf{u}_p^{(n)T}, n = 1, \dots, N\}_{p=1}^P$. The TVP from \mathcal{X} to \mathbf{y} is then

$$\mathbf{y} = \mathcal{X} \times_{n=1}^N \{\mathbf{u}_p^{(n)T}, n = 1, \dots, N\}_{p=1}^P, \quad (13)$$



(a) Tensor-to-vector projection.



(b) Elementary multilinear projection.

Fig. 6. Illustration of tensor-to-vector projection through elementary multilinear projection (EMP).

where the p th component of \mathbf{y} is obtained from the p th EMP as:

$$\mathbf{y}(p) = \mathcal{X} \times_1 \mathbf{u}_p^{(1)T} \times_2 \mathbf{u}_p^{(2)T} \dots \times_N \mathbf{u}_p^{(N)T}. \quad (14)$$

2.4 Relationships between Multilinear Projections

The following examines the relationships between the three basic multilinear projections. It is easy to verify that VVP is the special case of TTP and TVP with $N = 1$, and EMP is a degenerated version of TTP with $P_n = 1$ for all n . On the other hand, each projected element in TTP can be viewed as the projection of an EMP formed by taking one column from each of the

projection matrices. Thus, the projected tensor in TTP is effectively obtained through $\prod_{n=1}^N P_n$ interdependent EMPs, while in TVP, the P EMPs obtained sequentially are not interdependent in general.

Furthermore, the projection using an EMP $\{\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T}, \dots, \mathbf{u}^{(N)T}\}$ can be written as [28]

$$y = \langle \mathcal{X}, \mathcal{U} \rangle = \langle \text{vec}(\mathcal{X}), \text{vec}(\mathcal{U}) \rangle = [\text{vec}(\mathcal{U})]^T \text{vec}(\mathcal{X}). \quad (15)$$

Thus, an EMP is equivalent to a linear projection of $\text{vec}(\mathcal{X})$, the vectorized representation of \mathcal{X} , on a vector $\text{vec}(\mathcal{U})$. Since $\mathcal{U} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)}$, (15) indicates that EMP is equivalent to a linear projection with constraint on the projection vector such that it is the vectorized representation of a rank-one tensor.

The number of parameters to be estimated in a particular projection is an important concern in practice. Compared with a projection vector of size $I \times 1$ in VVP specified by I parameters ($I = \prod_{n=1}^N I_n$ for an N th-order tensor), an EMP in TVP can be specified by $\sum_{n=1}^N I_n$ parameters. Hence, to project a tensor of size $\prod_{n=1}^N I_n$ to a vector of size $P \times 1$, TVP needs to estimate only $P \cdot \sum_{n=1}^N I_n$ parameters, while VVP needs to estimate $P \cdot \prod_{n=1}^N I_n$ parameters. The implication in pattern recognition problem is that TVP has fewer parameters to estimate while being more constrained on the solutions, and VVP has less constraint on the solutions sought while having more parameters to estimate. For TTP with the same amount of dimensionality reduction $\prod_{n=1}^N P_n = P$, $\sum_{n=1}^N P_n \times I_n$ parameters need to be estimated. Table 3 contrasts the number of parameters to be estimated by the three projections for the same amount of dimensionality reduction. From the table, it can be seen that the parameter estimation problem in conventional VVP becomes extremely difficult for higher-order tensors. This often leads to the small sample size (SSS) problem in practice when there are limited number of training samples available.

Table 3

Number of parameters to be estimated by three multilinear projections.

Input	Output	VVP	TVP	TTP
$\prod_{n=1}^N I_n$	P	$P \cdot \prod_{n=1}^N I_n$	$P \cdot \sum_{n=1}^N I_n$	$\sum_{n=1}^N P_n \times I_n$
10×10	4	400	80	40 ($P_n = 2$)
100×100	4	40,000	800	400 ($P_n = 2$)
$100 \times 100 \times 100$	8	8,000,000	2,400	600 ($P_n = 2$)
$\prod_{n=1}^4 100$	16	1,600,000,000	6,400	800 ($P_n = 2$)

3 The Multilinear Subspace Learning Framework

This section formulates a general MSL framework. It defines the MSL problem in a similar way as LSL, as well as tensor and scalar scatter measures for optimality criterion construction. It also outlines a typical solution and discusses related issues.

3.1 Linear Subspace Learning

LSL algorithms [18, 44] solve for a linear projection satisfying some optimality criteria, given a set of training samples. The problem can be formulated as follows.

Linear Subspace Learning: A set of M vectorial samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ is available for learning, where each sample \mathbf{x}_m is an $I \times 1$ vector in a vector space \mathbb{R}^I . The LSL objective is to find a linear transformation (projection) $\mathbf{U} \in \mathbb{R}^{I \times P}$ such that the projected samples (the extracted features) $\{\mathbf{y}_m = \mathbf{U}^T \mathbf{x}_m\}$ satisfy an optimality criterion, where $\mathbf{y}_m \in \mathbb{R}^{P \times 1}$ and $P < I$.

Among various LSL algorithms, principal component analysis (PCA) [19] and linear discriminant analysis (LDA) [44] are two most widely used ones in a broad range of applications [49, 50]. PCA is an unsupervised algorithm that does not require labels for the training samples, while LDA is a supervised method that makes use of class specific information. Other popular LSL algorithms include independent component analysis (ICA) [51] and canonical correlation analysis (CCA) [52].

3.2 Multilinear Subspace Learning

MSL is the multilinear extension of LSL. It solves for a multilinear projection with some optimality criteria, given a set of training samples. This problem can be formulated similarly as follows.

Multilinear Subspace Learning: A set of M N th-order tensorial samples $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_M\}$ is available for learning, where each sample \mathcal{X}_m is an $I_1 \times I_2 \times \dots \times I_N$ tensor in a tensor space $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. The MSL objective is to find a multilinear transformation (projection) such that the projected samples (the extracted features) satisfy an optimality criterion, where the dimensionality of the projected space is much lower than the original tensor space.

Mathematically, the MSL problem can be written in a general form as

$$\{\mathbf{U}^{(n)}\} = \arg \max_{\{\mathbf{U}^{(n)}\}} \Phi(\{\mathbf{U}^{(n)}\}, \{\mathcal{X}_m\}) \quad (16)$$

or

$$\{\mathbf{u}_p^{(n)T}\}_{p=1}^P = \arg \max_{\{\mathbf{u}_p^{(n)T}\}_{p=1}^P} \Phi(\{\mathbf{u}_p^{(n)T}\}_{p=1}^P, \{\mathcal{X}_m\}), \quad (17)$$

where $\Phi(\cdot)$ denotes a criterion function to be maximized, or without loss of generality, $-\Phi(\cdot)$ is to be minimized. $\{\mathbf{U}^{(n)}\}$, $\{\mathbf{u}_p^{(n)T}\}_{p=1}^P$, and $\{\mathcal{X}_m\}$ are more compact forms of $\{\mathbf{U}^{(n)}, n = 1, \dots, N\}$, $\{\mathbf{u}_p^{(n)T}, n = 1, \dots, N\}_{p=1}^P$, and $\{\mathcal{X}_m, m = 1, \dots, M\}$, respectively.

Two key components for MSL are the multilinear projection employed and the objective criterion to be optimized. The projection to be solved can be any of the three types of basic multilinear projections discussed in Sec. 2.3. Thus, the well-studied LSL can be viewed as a special case of MSL where the projection to be solved is a VVP. Thus, the focus of this paper will be on MSL through TTP and TVP. This general formulation of MSL is important for evaluating, comparing, and further developing MSL solutions.

3.3 Scatter Measures for Multilinear Subspace Learning

In analogy to the definition of various scatters for vectorial features in LSL [44], tensor-based and scalar-based scatters in MSL are defined here.

Definition 1 Let $\{\mathcal{A}_m, m = 1, \dots, M\}$ be a set of M tensor samples in $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$. The total scatter of these tensors is defined as:

$$\Psi_{T_A} = \sum_{m=1}^M \|\mathcal{A}_m - \bar{\mathcal{A}}\|_F^2, \quad (18)$$

where $\bar{\mathcal{A}}$ is the mean tensor calculated as

$$\bar{\mathcal{A}} = \frac{1}{M} \sum_{m=1}^M \mathcal{A}_m. \quad (19)$$

The n -mode total scatter matrix of these samples is then defined as:

$$\mathbf{S}_{T_A}^{(n)} = \sum_{m=1}^M (\mathbf{A}_{m(n)} - \bar{\mathbf{A}}_{(n)}) (\mathbf{A}_{m(n)} - \bar{\mathbf{A}}_{(n)})^T, \quad (20)$$

where $\mathbf{A}_{m(n)}$ is the n -mode unfolded matrix of \mathcal{A}_m .

Definition 2 Let $\{\mathcal{A}_m, m = 1, \dots, M\}$ be a set of M tensor samples in $\mathbb{R}^{I_1} \otimes$

$\mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$. The between-class scatter and the within-class scatter of these tensors are defined as:

$$\Psi_{B_{\mathcal{A}}} = \sum_{c=1}^C M_c \|\bar{\mathcal{A}}_c - \bar{\mathcal{A}}\|_F^2, \text{ and } \Psi_{W_{\mathcal{A}}} = \sum_{m=1}^M \|\mathcal{A}_m - \bar{\mathcal{A}}_{c_m}\|_F^2, \quad (21)$$

respectively, where C is the number of classes, M_c is the number of samples for class c , c_m is the class label for the m th sample \mathcal{A}_m , $\bar{\mathcal{A}}$ is the mean tensor, and the class mean tensor is

$$\bar{\mathcal{A}}_c = \frac{1}{M_c} \sum_{m, c_m=c} \mathcal{A}_m. \quad (22)$$

Next, the n -mode scatter matrices are defined accordingly.

Definition 3 The n -mode between-class scatter matrix of these samples is defined as:

$$\mathbf{S}_{B_{\mathcal{A}}}^{(n)} = \sum_{c=1}^C M_c \cdot (\bar{\mathbf{A}}_{c(n)} - \bar{\mathbf{A}}_{(n)}) (\bar{\mathbf{A}}_{c(n)} - \bar{\mathbf{A}}_{(n)})^T, \quad (23)$$

and the n -mode within-class scatter matrix of these samples is defined as:

$$\mathbf{S}_{W_{\mathcal{A}}}^{(n)} = \sum_{m=1}^M (\mathbf{A}_{m(n)} - \bar{\mathbf{A}}_{c_m(n)}) (\mathbf{A}_{m(n)} - \bar{\mathbf{A}}_{c_m(n)})^T, \quad (24)$$

where $\bar{\mathbf{A}}_{c(n)}$ is the n -mode unfolded matrix of $\bar{\mathcal{A}}_c$.

The tensor scatters defined above are for MSL based on TTP. For MSL based on TVP, scalar-based scatters are defined, which can be viewed as degenerated versions of the vector-based or tensor-based scatters.

Definition 4 Let $\{a_m, m = 1, \dots, M\}$ be a set of M scalar samples. The total scatter of these scalars is defined as:

$$S_{T_{\mathbf{a}}} = \sum_{m=1}^M (a_m - \bar{a})^2, \quad (25)$$

where \bar{a} is the mean scalar calculated as

$$\bar{a} = \frac{1}{M} \sum_{m=1}^M a_m. \quad (26)$$

Definition 5 Let $\{a_m, m = 1, \dots, M\}$ be a set of M scalar samples. The

between-class scatter of these scalars is defined as:

$$S_{B_{\mathbf{a}}} = \sum_{c=1}^C M_c (\bar{a}_c - \bar{a})^2, \quad (27)$$

and the within-class scatter of these scalars is defined as:

$$S_{W_{\mathbf{a}}} = \sum_{m=1}^M (a_m - \bar{a}_{c_m})^2, \quad (28)$$

where

$$\bar{a}_c = \frac{1}{M_c} \sum_{m, c_m=c} a_m. \quad (29)$$

3.4 Typical Approach and Algorithmic Issues

While a linear projection (VVP) in LSL often has closed-form solutions, this is not the case for TTP and TVP in MSL. Instead, these two tensor-based projections have N sets of parameters to be solved, one in each mode, and the solution to one set often depends on the other sets (except when $N = 1$, the linear case), making their simultaneous estimation extremely difficult, if not impossible. Therefore, a suboptimal, iterative procedure originated from the alternating least square (ALS) algorithm [53–55] is usually employed to solve the tensor-based projections by alternating between solving one set of parameters (in one mode) at a time. Specifically, the parameters for each mode are estimated in turn separately and are conditioned on the parameter values for the other modes. At each step, by fixing the parameters in all the modes but one mode, a new objective function depending only on the mode left free to vary is optimized and this conditional subproblem is linear and much simpler to solve through unfolding tensors to matrices. The parameter estimations for each mode are obtained in this way sequentially and iteratively until convergence. This process is described in Fig. 7(a) and also illustrated in Fig. 7(b).

Consequently, the issues due to the iterative nature of the solution, such as initialization, the order of solving the projections, termination, and convergence, need to be addressed. In addition, for MSL through TTP, a mechanism is often needed to determine the desired subspace dimensionality $\{P_1, P_2, \dots, P_N\}$. This is because it is costly to exhaustively test the large number of possible combinations of the N values, P_1, P_2, \dots, P_N , for a specific amount of dimensionality reduction, especially for higher-order tensors. In contrast, the value P is relatively easier to determine for MSL through TVP.

To end this section, Table 4 summarizes the key differences between MSL

Input: A set of tensor samples $\{\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}, m = 1, \dots, M\}$, the desired tensor subspace dimensionality.

Output: The multilinear projection that maximizes an optimality criterion in the projected space.

Algorithm:

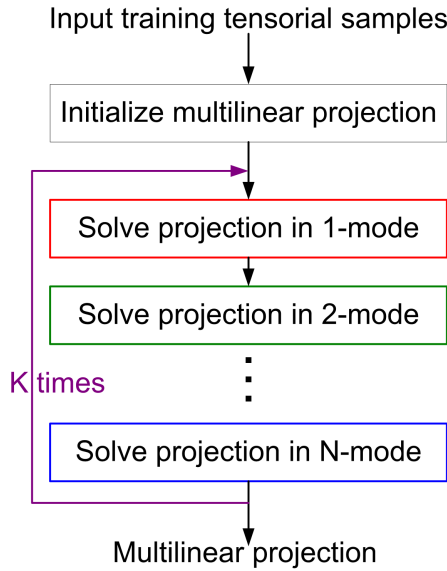
Step 1 (Initialization): Initialize the multilinear projection.

Step 2 (Local optimization):

- For $k = 1 : K$
 - For $n = 1 : N$

Solve for the n -mode projection as a linear problem obtained through fixing projections in all the other modes and unfolding tensors to matrices.
 - If $k = K$ or the algorithm converges, break and output the current multilinear projection.
-

(a) The pseudo-code.



(b) Algorithm flow.

Fig. 7. Illustration of a typical multilinear subspace learning algorithm.

and LSL. In the table, massive data refers to data with its dimensionality beyond the processing power of common computational hardwares when LSL algorithms are used, such as high-resolution face images or standard gait silhouette sequences. A key difference between MSL and LSL is the number of parameters to be estimated for the same dimensionality reduction problem, as illustrated in Table 3. Consequently, as pointed out in [22], MSL is less susceptible to the SSS problem and overfitting compared to LSL.

Table 4
Linear versus multilinear subspace learning.

Comparison	Linear subspace learning	Multilinear subspace learning
<i>Representation</i>	Reshape into vectors	Natural tensorial representation
<i>Structure</i>	Break natural structure	Preserve natural structure
<i>Parameter</i>	Estimate a large number of parameters	Estimate fewer parameters
<i>SSS problem</i>	More severe SSS problem	Less SSS problem
<i>Massive data</i>	Hardly applicable to massive data	Able to handle massive data
<i>Optimization</i>	Closed-form solution	Suboptimal, iterative solution

4 Taxonomy of Multilinear Subspace Learning Algorithms

This section reviews several important MSL algorithms under the MSL framework. Due to the fundamentality and importance of PCA and LDA, the focus is on the multilinear extensions of these two classical linear algorithms. Figures 8 and 9 depict taxonomies for these algorithms, one for multilinear extensions of PCA and the other for multilinear extensions of LDA, respectively, which will be discussed in detail in the following. The historical basis of MSL will also be described.

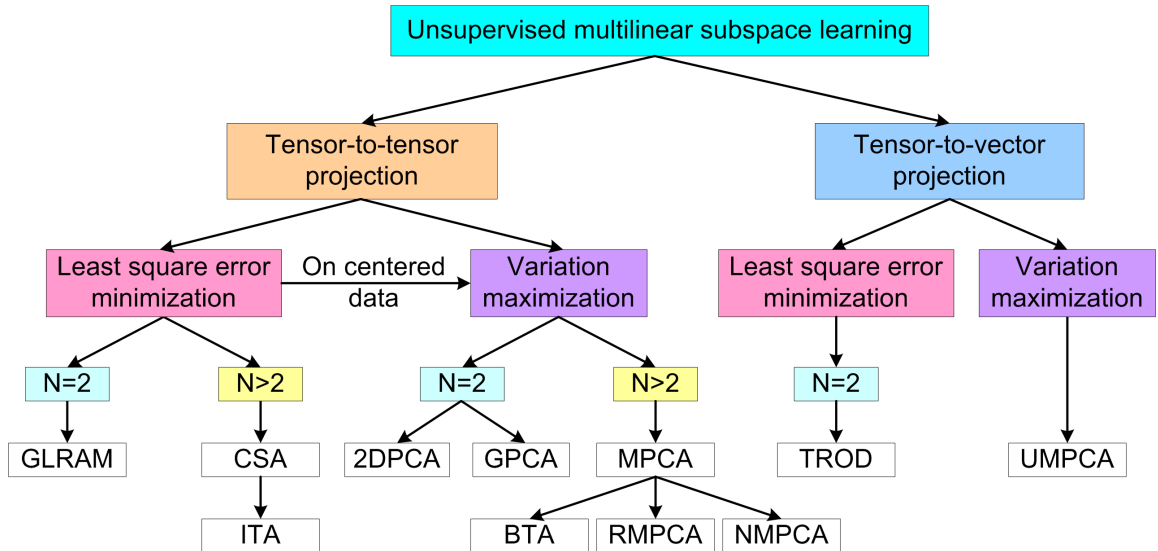


Fig. 8. A taxonomy of unsupervised multilinear subspace learning algorithms.

4.1 Unsupervised Multilinear Subspace Learning Algorithms

The development of unsupervised MSL started with the treatment of images directly as matrices rather than vectors.

4.1.1 Unsupervised MSL through TTP

A two-dimensional PCA (2DPCA) algorithm is proposed in [56]. This algorithm solves for a linear transformation that projects an image to a low-dimensional matrix while maximizing the variance measure. It works directly on image matrices but there is only one linear transformation in the 2-mode. Thus, the image data is projected in the 2-mode (the row mode) only while the projection in the 1-mode (the column mode) is ignored, resulting in poor dimensionality reduction. The projection can be viewed as a special case of second-order TTP, where the 1-mode projection matrix is an identity matrix.

A more general algorithm named the generalized low rank approximation of matrices (GLRAM) was introduced in [57], which takes into account the spatial correlation of the image pixels within a localized neighborhood and applies two linear transforms to both the left and right sides of input image matrices. This algorithm solves for two linear transformations that project an image to a low-dimensional matrix while minimizing the least-square (reconstruction) error measure. Thus, projections in both modes are involved and the projection is TTP with $N = 2$. Better dimensionality reduction results than [56] are obtained according to [57].

Although GLRAM exploits both modes for subspace learning, it is formulated for matrices only. Later, the work in [58] presents tensor approximation methods for third-order tensors using slice projection, while the so-called concur-

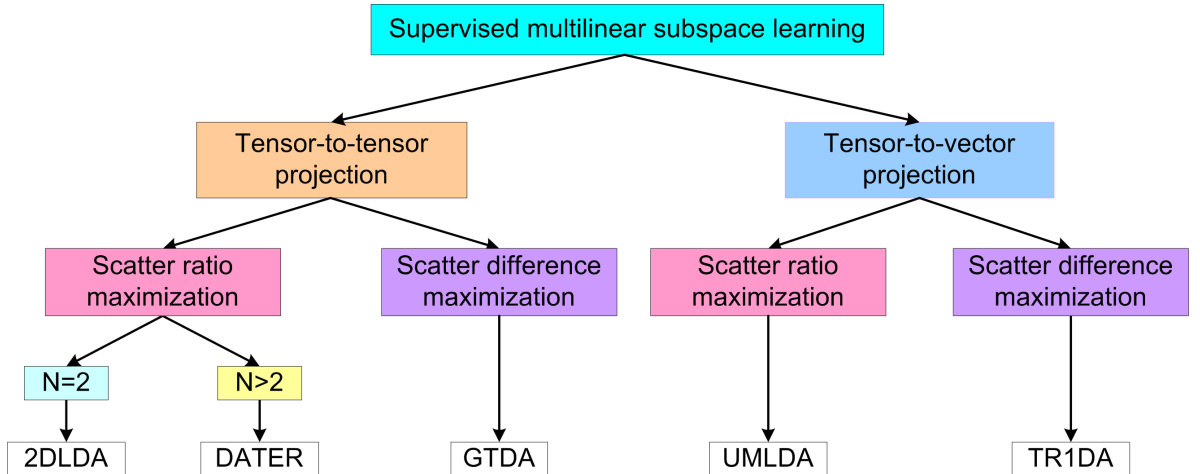


Fig. 9. A taxonomy of supervised multilinear subspace learning algorithms.

rent subspaces analysis (CSA) is formulated in [27] for general tensor objects as a generalization of GLRAM for higher-order tensors. The CSA algorithm solves for a TTP minimizing a reconstruction error metric, however, how to determine the tensor subspace dimensionality is not addressed in this work.

Whereas GLRAM and CSA advanced unsupervised MSL, they are both formulated with the objective of optimal reconstruction or approximation of tensors. Therefore, they ignored an important centering step in unsupervised subspace learning algorithms developed for recognition, such as the classical PCA, where the data is centered first before obtaining the subspace projection. It should be pointed out that for the reconstruction or approximation problem, centering is not essential, as the (sample) mean is the main focus of attention [21]. However, in recognition applications where the solutions involve eigenproblems, non-centering (in other words, an average different from zero) can potentially affect the eigen-decomposition in each mode and lead to a solution that captures the variation with respect to the origin rather than capturing the true variation of the data (with respect to the data center) [21].

In contrast, the generalized PCA (GPCA) proposed in [1] is an extension of PCA that works on matrices. GPCA is exactly the same as GLRAM except that the projection takes the centered data rather than the original coordinate as input. Nonetheless, this work is formulated only for matrices, and important issues such as initialization and subspace dimensionality determination are not studied in this work either.

The multilinear PCA (MPCA) algorithm proposed in [21] generalizes GPCA to work for tensors of any order, where the objective is to find a TTP that captures most of the original tensorial input variations (Definition 1). Furthermore, methods for systematic subspace dimensionality determination are proposed for the first time in the literature, in contrast to heuristic methods in [27, 57]. With the introduction of a discriminative tensor feature selection mechanism, MPCA is further combined with LDA for general tensor object recognition in [21].

In [59], a TTP-based MSL algorithm named Bayesian tensor analysis (BTA) is proposed to generalize the Bayesian PCA algorithm [60] and develop probabilistic graphical models for tensors. BTA is able to determine the subspace dimensionality automatically by setting a series of hyper parameters. As the mean is subtracted from the input tensor in dimensionality reduction using BTA, it can be considered as a probabilistic extension of MPCA as well. In [61], two robust MPCA (RMPCA) algorithms are proposed, where iterative algorithms are derived on the basis of Lagrange multipliers to deal with sample outliers and intra-sample outliers. In [62], the non-negative MPCA (NMPCA) extends MPCA to constrain the projection matrices to be non-negative. Furthermore, NMPCA preserves the non-negativity of the original tensor samples,

which is important when the underlying data factors have physical or psychological interpretation. The solution for NMPCA is developed by exploiting the structure of the Grassmann manifold [62].

In addition, an incremental tensor analysis (ITA) framework is proposed in [17] for summarizing higher-order data streams represented as tensors. The data summary is obtained through TTP and is updated incrementally. Three variants of ITA are introduced in [17]: dynamic tensor analysis (DTA) that incrementally maintains covariance matrices for all modes and uses the leading eigenvectors of covariance matrices as projection matrices [15], streaming tensor analysis (STA) that directly updates the leading eigenvectors of covariance matrices using the SPIRIT algorithm [63], and window-based tensor analysis (WTA) that uses similar updates as DTA while performing alternating iteration to further improve the results. The ITA framework focuses on approximation problem, hence, the objective is to minimize the least square error and it can be considered as an incremental version of CSA for streaming data.

4.1.2 *Unsupervised MSL through TVP*

In comparison to the review above, there are much fewer unsupervised MSL algorithms based on TVP. The tensor rank-one decomposition (TROD) algorithm introduced in [64] is TVP-based and it is formulated only for image matrices. This algorithm looks for a second-order TVP that projects an image to a low-dimensional vector while minimizing a least-square (reconstruction) error measure. Hence, the input data is not centered before learning. The solution of TROD relies on a heuristic procedure of successive residue calculation, i.e., after obtaining the p th EMP, the input image is replaced by its residue.

None of the above unsupervised MSL algorithms takes into account the correlations among features and shares an important property with PCA, i.e., zero-correlation among extracted features. It is well-known that PCA derives uncorrelated features, which contain minimum redundancy and ensure linear independence among features. Uncorrelated features can also greatly simplify the subsequent classification task and they are highly desirable in recognition applications. An uncorrelated MPCA (UMPCA) algorithm is proposed in [25], which extracts uncorrelated multilinear features through TVP while capturing most of the variation in the original data input (Definition 4). The UMPCA solution consists of sequential iterative steps for successive variance maximization. The work in [25] has also derived a systematic way to determine the maximum number of uncorrelated multilinear features that can be extracted by the method.

4.2 Supervised Multilinear Subspace Learning Algorithms

Similar to unsupervised MSL, the development of supervised MSL started with 2D extensions of LDA.

4.2.1 Supervised MSL through TTP

Like GLRAM and GPCA, the 2D LDA (2DLDA) introduced in [65] solves for two linear transformations that project an image to a low-dimensional matrix, but with a different objective criterion. For the input image samples, the between-class and within-class scatter measures are defined for matrix representations (Definition 2). A matrix-based discrimination criterion is then defined as the scatter ratio, which is to be maximized in 2DLDA. Unlike the unsupervised MSL algorithms reviewed above, 2DLDA does not converge over iterations.

Later, as a higher-order extension of 2DLDA, the discriminant analysis with tensor representation (DATER)¹ was proposed to perform discriminant analysis on more general tensorial inputs [2]. The DATER algorithm solves for a TTP maximizing the tensor-based scatter ratio (Definition 2). However, this algorithm does not converge over iterations either.

In [23], the general tensor discriminant analysis (GTDA) algorithm is proposed. The GTDA algorithm also solves for a TTP. The difference with DATER is that it maximizes a tensor-based scatter difference criterion (Definition 2), with a tuning parameter involved [66]. The criterion used is dependent on the coordinate system, as pointed out in [67], and the tuning parameter is heuristically determined in [23]. In contrast with 2DLDA/DATER, this algorithm has good convergence property [23] and it is the first discriminative MSL algorithm that converges to a local solution.

4.2.2 Supervised MSL through TVP

In this category, the first algorithm is the tensor rank-one discriminant analysis (TR1DA) algorithm proposed in [46, 47], derived from the TROD algorithm [64]. The TR1DA algorithm is formulated for general tensor objects and it looks for a TVP that projects a tensor to a low-dimensional vector while maximizing the scalar scatter difference criterion (Definition 5). Therefore, the criterion is also dependent on the coordinate system and there is no way to determine the optimal tuning parameter either. Furthermore, as in TROD, this

¹ Here, the name used when the algorithm was first proposed is adopted as it is more commonly referred to in the literature.

algorithm also relies on the repeatedly-calculated residues, originally proposed in [68] for tensor approximation. The adoption of this heuristic procedure here lacks theoretical explanation for a discriminative criterion.

Similar to the case of unsupervised MSL, the supervised MSL algorithms discussed so far do not take the correlations among features into account and they do not derive uncorrelated features as in the classical LDA [69, 70]. As mentioned in Sec. 4.1.2, uncorrelated features are highly desirable in many applications [70]. An uncorrelated multilinear discriminant analysis (UMLDA) algorithm is formulated in [26]. UMLDA aims to extract uncorrelated discriminative features directly from tensorial data through solving a TVP to maximize a scalar scatter ratio criterion (Definition 5). The solution consists of sequential iterative processes and incorporates an adaptive regularization procedure to enhance the performance in the small sample size scenario. Furthermore, an aggregation scheme is adopted to combine differently initialized and regularized UMLDA recognizers for enhanced generalization performance while alleviating the regularization parameter selection problem. This extension is called regularized UMLDA with aggregation (R-UMLDA-A) [26].

4.3 *The Historical Basis of MSL*

Multilinear algebra, the extension of linear algebra, has been studied in mathematics around the middle of the 20th century [29]. It built on the concept of tensors and developed the theory of tensor spaces.

A popular early application of multilinear algebra is multi-way analysis in psychometrics and chemometrics starting from the 60s and 70s, for factor analysis of multi-way data sets, which are higher-order tensors characterized by several sets of categorical variables that are measured in a crossed fashion [53–55, 71]. Two main types of decomposition methods have been developed in this field: the Tucker decomposition [30, 41, 71, 72], and the canonical decomposition (CANDECOMP) [30, 41, 53], also known as the parallel factors (PARAFAC) decomposition [30, 41, 54]. There are also other tensor decompositions such as CANDECOMP with linear constraints (CANDELINC) [73].

In the 90s, the developments in higher-order statistics of multivariate stochastic variables have attracted interests in higher-order tensors from the signal processing community [74]. The Tucker decomposition was reintroduced and further developed in [30] as an extension of the singular value decomposition (SVD) to higher-order tensors: the higher-order SVD (HOSVD) solution. Its computation leads to the calculation of N different matrix SVDs of differently unfolded matrices. The ALS algorithm for the best Rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors was studied in [31], where tensor data was

iteratively projected into a lower dimensional tensor space. The application of HOSVD truncation and the best Rank- (R_1, R_2, \dots, R_N) approximation to dimensionality reduction in ICA was discussed in [75].

The work in [30,31] led to the development of new multilinear algorithms and the exploration of new application areas for tensor data analysis. Multilinear analysis of image data is pioneered by the TensorFace method [76,77], which employs the multilinear algorithms proposed in [30,31] to analyze the factors involved in the formation of facial images. Similar analysis has also been done for motion signatures [78] and gait sequences [79]. However, in these multiple factor analysis work, input data such as images or video sequences are still represented as vectors. These vectors are arranged into a tensor according to multiple factors involved in their formation for subsequent analysis. Such tensor formation needs a large number of training samples captured under various conditions, which is often impractical and may have the missing-data problem. Furthermore, the tensor data size is usually huge, leading to high memory and computational demands.

In the last few years, several methods were proposed for direct learning of a subspace from tensorial data [1,2,21,23,25–27,47,56,64,65]. Besides the MSL algorithms reviewed above, there is also a multilinear extension of the CCA algorithm named as tensor CCA in [80]. In addition, solutions are proposed in [81,82] to rearrange elements within a tensor to maximize the correlations among n -mode vectors for better dimensionality reduction performance. Furthermore, systematic treatment of this topic has appeared in [28,42]. Besides the multilinear extensions of LSL algorithms, multilinear extensions of linear graph-embedding algorithms were also introduced in [43,48,83–86], in a similar fashion as the existing MSL algorithms reviewed in this paper.

5 Multilinear Subspace Learning Applications

Due to the advances in sensor and storage technology, MSL is becoming increasingly popular in a wide range of application domains involving tensor-structured data sets. This section will summarize several applications of MSL algorithms in real-world applications, including face recognition and gait recognition in biometrics, music genre classification in audio signal processing, EEG signal classification in biomedical engineering, anomaly detection in data mining, and visual content analysis in computer vision. Other MSL applications include handwritten digit recognition [27,57], image compression [1,27,64], image/video retrieval [1,87], and object categorization and recognition [46]. For more general tensor data applications, [35], [88], [89] and [90] are good references.

5.1 Face Recognition and Gait Recognition

Face and gait are two typical physiological and behavioral biometrics, respectively. Compared with other biometric traits, face and gait have the unique property that they facilitate human recognition at a distance, which is extremely important in surveillance applications. Face recognition has a large number of commercial security and forensic applications, including video surveillance, access control, mugshot identification, and video communications [10, 91]. Gait is a person’s walking style and it is a complex spatio-temporal biometric [10, 91]. The interest in gait recognition is strongly motivated by the need for an automated human identification system at a distance in visual surveillance and monitoring applications in security-sensitive environments, e.g., banks, parking lots, malls, and transportation hubs such as airports and train stations [21]. Many MSL algorithms are first applied to appearance-based learning for face recognition [2, 27, 56, 57, 86, 92, 93] and/or gait recognition [21, 23, 94–97], where the input face images or binary gait silhouette sequences are treated as tensorial holistic patterns as shown in Fig. 1.

5.2 Music Genre Classification

Music genre is a popular description of music content for music database organization. The NMPCA algorithm proposed in [62] is designed for music genre classification, combined with the nearest neighbor or support vector machines (SVM) classifiers. A 2D joint acoustic and modulation frequency representation of music signals is utilized in this work to capture slow temporal modulations [62], inspired by psychophysiological investigations on the human auditory system. The acoustic signal is first converted to a time-frequency auditory spectrogram and a wavelet transform is applied to each row of this spectrogram to estimate its temporal modulation content through modulation scale analysis [62]. In three different sets of experiments on public music genre data sets, NMPCA has achieved the state-of-the-art classification results.

5.3 EEG Signal Classification

Electroencephalography (EEG) records brain activities as multichannel time series from multiple electrodes placed on the scalp of a subject to provide a direct communication channel between brain and computer, and it is widely used in noninvasive brain computer interfaces (BCI) applications [98]. In [5], a tensor-based EEG classification scheme is proposed, where the wavelet transform is applied to EEG signals to result in third-order tensor representations

in the spatial-spectral-temporal domain. GTDA is then applied to obtain low-dimensional tensors, from which discriminative features are selected for SVM classification. Motor imagery experiments on three data sets demonstrate that the proposed scheme can outperform many other existing EEG signal classification schemes, especially when there is no prior neurophysiologic knowledge available for EEG signal preprocessing.

5.4 *Anomaly Detection*

In streaming data applications such as network forensics, large volumes of high-order data continuously arrive incrementally. It is challenging to perform incremental pattern discovery on such data streams. The ITA framework in [17] is devised to tackle this problem and anomaly detection is one of its targeted applications. The abnormality is modeled by the reconstruction error of incoming tensor data streams and a large reconstruction error often indicates an anomaly. The proposed method is illustrated on a network monitoring example, where three types of anomalies (abnormal source hosts, abnormal destination hosts, and abnormal ports) can be detected with very high precision.

5.5 *Visual Content Analysis*

As a dimensionality reduction technique, MSL can also be used for various visual content analysis tasks [87, 99–103]. E.g., the MPCA algorithm is used to visualize and summarize a crowd video sequence in a 2D subspace in [100]. In [101], an optical flow tensor is built and GTDA is used to reduce the dimensionality for further semantic analysis of video sequences. For 3D facial data modeling, the BTA algorithm is employed for the 3D facial expression retargeting task [59]. In [102], a modified version of the DTA algorithm (with mean and variance updating) is applied on a weighted tensor representation for visual tracking of human faces, and in [103], an automatic human age estimation method is developed using MPCA for dimensionality reduction of tensorial Gaussian receptive maps.

6 **Conclusions and Future Works**

This paper presents a survey of an emerging dimensionality reduction approach for direct feature extraction from tensor data: multilinear subspace learning. It reduces the dimensionality of massive data directly from their

natural multidimensional representation: tensors. This survey covers multilinear projections, MSL framework, typical MSL solutions, MSL taxonomies, and MSL applications. MSL is a new field with many open issues to be examined further. The rest of this section outlines several research topics that worth further investigation. Two main directions have been identified. One is towards the development of new MSL solutions, while the other is towards the exploration of new MSL applications.

In future research, new algorithms can be investigated along the following directions. The systematic treatment on MSL will benefit the development of new multilinear learning algorithms, especially by extending the rich ideas and algorithms in the linear counterparts to the multilinear case. This paper has focused on the extensions of PCA and LDA to their multilinear counterparts. In future work, more complex multilinear mapping can be investigated, for example, by developing multilinear extensions of graph-embedding algorithms such as Isomap [104], Locally Linear Embedding [105], and Locality Preserving Projections [106,107] under the MSL framework. As mentioned in Section 4.3, there have been some developments already in this area [43,48,84–86]. The MSL framework can help the understanding of these existing solutions and it can also benefit their further development. Furthermore, in MSL, there are still many open problems remaining, such as the optimal initialization, the optimal projection order, and the optimal stopping criterion. There has been some attempts in solving some of these problems in [21]. However, further research in this direction is needed for deeper understanding on these issues and even alternative optimization strategies can be explored.

Besides the applications reviewed in Sec. 5, there are a wide range of applications dealing with real-world tensor objects where MSL may be useful, as mentioned in the Introduction of this paper. Examples include high-resolution and 3D face detection and recognition [108–111], clustering or retrieval of images or 3D objects [8,112], space-time analysis of video sequences [11,113], and space-time super-resolution [14]. In addition, massive streaming data are frequently organized as multidimensional objects, such as those in social network analysis, web data mining, and sensor network analysis [17]. Some tensor-based techniques have been developed in these fields [15,16] and further investigation of MSL for these applications under the framework presented in this paper can be fruitful.

To summarize, the recent prevalence of applications involving massive multidimensional data has increased the demand for technical developments in the emerging research field of MSL. This overview provides a foundation upon which solutions for many interesting and challenging problems in tensor data applications can be built. It is the authors’ hope that this unifying review and the insights provided in this paper will foster more principled and successful applications of MSL in a wide range of research disciplines.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful comments, which have helped to improve the quality of this paper.

References

- [1] J. Ye, R. Janardan, Q. Li, GPCA: An efficient dimension reduction scheme for image compression and retrieval, in: The Tenth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2004, pp. 354–363.
- [2] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H. Zhang, Multilinear discriminant analysis for face recognition, *IEEE Transactions on Image Processing* 16 (1) (2007) 212–220.
- [3] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, *IEEE Transactions on Neural Networks* 14 (1) (2003) 117–126.
- [4] H. Lu, J. Wang, K. N. Plataniotis, A review on face and gait recognition: System, data and algorithms, in: S. Stergiopoulos (Ed.), *Advanced Signal Processing Handbook*, 2nd Edition, CRC Press, Boca Raton, Florida, 2009, pp. 303–330.
- [5] J. Li, L. Zhang, D. Tao, H. Sun, Q. Zhao, A prior neurophysiologic knowledge free tensor-based scheme for single trial eeg classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 17 (2) (2009) 107–115.
- [6] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Regularized common spatial patterns with generic learning for EEG signal classification, in: *Proc. 31st Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, 2009.
- [7] J. Ye, T. Li, T. Xiong, R. Janardan, Using uncorrelated discriminant analysis for tissue classification with gene expression data, *IEEE/ACM Trans. Comput. Biology Bioinformatics* 1 (4) (2004) 181–190.
- [8] H. S. Sahambi, K. Khorasani, A neural-network appearance-based 3-D object recognition using independent component analysis, *IEEE Transactions on Neural Networks* 14 (1) (2003) 138–149.
- [9] N. Renard, S. Bourennane, Dimensionality reduction based on tensor modeling for classification methods, *IEEE Transactions on Geoscience and Remote Sensing* 47 (4) (2009) 1123–1131.
- [10] R. Chellappa, A. Roy-Chowdhury, S. Zhou, *Recognition of Humans and Their Activities Using Video*, Morgan & Claypool Publishers, San Rafael, California, 2005.

- [11] R. D. Green, L. Guan, Quantifying and recognizing human movement patterns from monocular video images-part II: applications to biometrics, *IEEE Transactions on Circuits and Systems for Video Technology* 14 (2) (2004) 191–198.
- [12] D. R. Hardoon, J. Shawe-Taylor, Decomposing the tensor kernel support vector machine for neuroscience data with structure labels, *Machine Learning* 79 (1-2) (2010) 29–46.
- [13] X. He, Incremental semi-supervised subspace learning for image retrieval, in: *ACM conference on Multimedia 2004*, 2004, pp. 2–8.
- [14] E. Shechtman, Y. Caspi, M. Irani, Space-time super-resolution, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (4) (2005) 531–545.
- [15] J. Sun, D. Tao, C. Faloutsos, Beyond streams and graphs: dynamic tensor analysis, in: *Proc. the 12th ACM SIGKDD int. conf. on Knowledge discovery and data mining*, 2006, pp. 374–383.
- [16] J. Sun, Y. Xie, H. Zhang, C. Faloutsos, Less is more: Sparse graph mining with compact matrix decomposition, *Statistical Analysis and Data Mining* 1 (1) (2008) 6–22.
- [17] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, C. Faloutsos, Incremental tensor analysis: Theory and applications, *ACM Trans. on Knowledge Discovery from Data* 2 (3) (2008) 11:1–11:37.
- [18] G. Shakhnarovich, B. Moghaddam, Face recognition in subspaces, in: S. Z. Li, A. K. Jain (Eds.), *Handbook of Face Recognition*, Springer-Verlag, 2004, pp. 141–168.
- [19] I. T. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer Series in Statistics, 2002.
- [20] S. Z. Li, A. K. Jain, Introduction, in: S. Z. Li, A. K. Jain (Eds.), *Handbook of Face Recognition*, Springer-Verlag, 2004, pp. 1–11.
- [21] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, MPCA: Multilinear principal component analysis of tensor objects, *IEEE Transactions on Neural Networks* 19 (1) (2008) 18–39.
- [22] D. Tao, X. Li, X. Wu, W. Hu, S. J. Maybank, Supervised tensor learning, *Knowledge and Information Systems* 13 (1) (2007) 1–42.
- [23] D. Tao, X. Li, X. Wu, S. J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (10) (2007) 1700–1715.
- [24] D. Tao, X. Li, X. Wu, S. J. Maybank, Tensor rank one discriminant analysis: a convergent method for discriminative multilinear subspace selection, *Neurocomputing* 71 (10-12) (2008) 1866–1882.

- [25] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning, *IEEE Transactions on Neural Networks* 20 (11) (2009) 1820–1836.
- [26] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition, *IEEE Transactions on Neural Networks* 20 (1) (2009) 103–123.
- [27] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, T. S. Huang, Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis, *IEEE Transactions on Circuits and Systems for Video Technology* 18 (1) (2008) 36–47.
- [28] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, A taxonomy of emerging multilinear discriminant analysis solutions for biometric signal recognition, in: N. V. Boulgouris, K. Plataniotis, E. Micheli-Tzanakou (Eds.), *Biometrics: Theory, Methods, and Applications*, Wiley/IEEE, 2009, pp. 21–45.
- [29] W. H. Greub, *Multilinear Algebra*, Springer-Verlag, Berlin, 1967.
- [30] L. D. Lathauwer, B. D. Moor, J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal of Matrix Analysis and Applications* 21 (4) (2000) 1253–1278.
- [31] L. D. Lathauwer, B. D. Moor, J. Vandewalle, On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors, *SIAM Journal of Matrix Analysis and Applications* 21 (4) (2000) 1324–1342.
- [32] L. Qi, W. Sun, Y. Wang, Numerical multilinear algebra and its applications, *FRONTIERS OF MATHEMATICS IN CHINA* 2 (4) (2007) 501–526.
- [33] D. Muti, S. Bourennane, Survey on tensor signal algebraic filtering, *Signal Processing* 87 (2) (2007) 237–249.
- [34] E. Acar, B. Yener, Unsupervised multiway data analysis: A literature survey, *IEEE Transactions on Knowledge and Data Engineering* 21 (1) (2009) 6–20.
- [35] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, *SIAM Review* 51 (3) (2009) 455–500.
- [36] S. Zafeiriou, Algorithms for nonnegative tensor factorization, in: S. Aja-Fernández, R. d. L. García, D. Tao, X. Li (Eds.), *Tensors in Image Processing and Computer Vision*, Springer, 2009, pp. 105–124.
- [37] T. Hazan, S. Polak, A. Shashua, Sparse image coding using a 3D non-negative tensor factorization, in: *Proc. IEEE Conference on Computer Vision*, Vol. 1, 2005, pp. 50–57.
- [38] A. Shashua, T. Hazan, Non-negative tensor factorization with applications to statistics and computer vision, in: *Proc. Int. Conf. on Machine Learning*, 2005, pp. 792–799.

- [39] R. Bro, Multi-way analysis in the food industry - models, algorithms and applications, Ph.D. thesis, University of Amsterdam, The Netherlands (1998). URL http://www.models.kvl.dk/sites/default/files/brothesis_0.pdf
- [40] A. K. Smilde, R. Bro, P. Geladi, Multi-way Analysis, John Wiley and Sons, 2004.
- [41] B. W. Bader, T. G. Kolda, Algorithm 862: Matlab tensor classes for fast algorithm prototyping, ACM Trans. on Mathematical Software 32 (4) (2006) 635–653.
- [42] H. Lu, Multilinear subspace learning for face and gait recognition, Ph.D. thesis, University of Toronto (2008). URL <https://tspace.library.utoronto.ca/handle/1807/16750>
- [43] X. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: Advances in Neural Information Processing Systems 18 (NIPS), 2005.
- [44] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, 2nd Edition, Wiley Interscience, 2001.
- [45] T. K. Moon, W. C. Stirling, Mathematical methods and Algorithms for Signal Processing, Prentice Hall, 2000.
- [46] Y. Wang, S. Gong, Tensor discriminant analysis for view-based object recognition, in: Proc. Int. Conf. on Pattern Recognition, Vol. 3, 2006, pp. 33–36.
- [47] D. Tao, X. Li, X. Wu, S. J. Maybank, Elapsed time in human gait recognition: A new approach, in: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 2, 2006, pp. 177–180.
- [48] G. Hua, P. A. Viola, S. M. Drucker, Face recognition using discriminatively trained orthogonal rank one tensor projections, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
- [49] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
- [50] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
- [51] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (1994) 287–314.
- [52] B. Thompson, Canonical correlation analysis: Uses and interpretation, Sage Publications, Thousand Oaks, CA, 1984.
- [53] J. D. Carroll, J. J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition, Psychometrika 35 (1970) 283–319.

- [54] R. A. Harshman, Foundations of the parafac procedure: Models and conditions for an “explanatory” multi-modal factor analysis, *UCLA Working Papers in Phonetics* 16 (1970) 1–84.
- [55] P. Kroonenberg, J. Leeuw, Principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika* 45 (1) (1980) 69–97.
- [56] J. Yang, D. Zhang, A. F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (1) (2004) 131–137.
- [57] J. Ye, Generalized low rank approximations of matrices, *Machine Learning* 61 (1-3) (2005) 167–191.
- [58] H. Wang, N. Ahuja, A tensor approximation approach to dimensionality reduction, *International Journal of Computer Vision* 76 (3) (2008) 217–229.
- [59] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, S. J. Maybank, Bayesian tensor approach for 3-D face modeling, *IEEE Transactions on Circuits and Systems for Video Technology* 18 (10) (2008) 1397–1410.
- [60] C. M. Bishop, Bayesian PCA, in: *Advances in Neural Information Processing Systems (NIPS)*, 1999, pp. 382–388.
- [61] K. Inoue, K. Hara, K. Urahama, Robust multilinear principal component analysis, in: *Proc. IEEE Conference on Computer Vision*, 2009, pp. 591–597.
- [62] Y. Panagakis, C. Kotropoulos, G. R. Arce, Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification, *IEEE Trans. on Audio, Speech, and Language Processing* 18 (3) (2010) 576–588.
- [63] S. Papadimitriou, J. Sun, C. Faloutsos, Streaming pattern discovery in multiple time-series, in: *Proc. 31st International Conference on Very Large Data Bases*, 2005, pp. 697–708.
- [64] A. Shashua, A. Levin, Linear image coding for regression and classification using the tensor-rank principle, in: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Vol. I, 2001, pp. 42–49.
- [65] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, in: *Advances in Neural Information Processing Systems (NIPS)*, 2004, pp. 1569–1576.
- [66] Q. Liu, X. Tang, H. Lu, S. Ma, Face recognition using kernel scatter-difference-based discriminant analysis, *IEEE Transactions on Neural Networks* 17 (4) (2006) 1081–1085.
- [67] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Boston, MA, 1990.

- [68] T. G. Kolda, Orthogonal tensor decompositions, *SIAM Journal of Matrix Analysis and Applications* 23 (1) (2001) 243–255.
- [69] Z. Jin, J. Y. Yang, Z. M. Tang, Z. S. Hu, A theorem on the uncorrelated optimal discriminant vectors, *Pattern Recognition* 34 (10) (2001) 2041–2047.
- [70] J. Ye, R. Janardan, Q. Li, H. Park, Feature reduction via generalized uncorrelated linear discriminant analysis, *IEEE Transactions on Knowledge and Data Engineering* 18 (10) (2006) 1312–1322.
- [71] L. R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika* 31 (1966) 279–311.
- [72] A. Kapteyn, H. Neudecker, T. Wansbeek, An approach to n-mode components analysis, *Psychometrika* 51 (1986) 269–275.
- [73] J. D. Carroll, S. Pruzansky, J. B. Kruskal, CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters, *Psychometrika* 45 (1980) 3–24.
- [74] P. Comon, B. Mourrain, Decomposition of quantics in sums of powers of linear forms, *Signal Processing* 53 (1996) 93–108.
- [75] L. D. Lathauwer, J. Vandewalle, Dimensionality reduction in higher-order signal processing and rank- (R_1, R_2, \dots, R_N) reduction in multilinear algebra, *Linear Algebra and its Applications* 391 (2004) 31–55.
- [76] M. A. O. Vasilescu, D. Terzopoulos, Multilinear analysis of image ensembles: Tensorfaces, in: *Proc. seventh European Conference on Computer Vision*, 2002, pp. 447–460.
- [77] M. A. O. Vasilescu, D. Terzopoulos, Multilinear image analysis for facial recognition, in: *Proc. Int. Conf. on Pattern Recognition*, Vol. 2, 2002, pp. 511–514.
- [78] M. A. O. Vasilescu, Human motion signatures: analysis, synthesis, recognition, in: *Proc. Int. Conf. on Pattern Recognition*, Vol. 3, 2002, pp. 456–460.
- [79] C. S. Lee, A. Elgammal, Towards scalable view-invariant gait recognition: Multilinear analysis for gait, in: *Proc. Int. Conf. on Audio and Video-Based Biometric Person Authentication*, 2005, pp. 395–405.
- [80] T.-K. Kim, R. Cipolla, Canonical correlation analysis of video volume tensors for action categorization and detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (8) (2009) 1415–1428.
- [81] D. Xu, S. Yan, S. Lin, T. S. Huang, S.-F. Chang, Enhancing bilinear subspace learning by element rearrangement, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (10) (2009) 1913–1920.
- [82] S. Yan, D. Xu, S. Lin, T. S. Huang, S.-F. Chang, Element rearrangement for tensor-based subspace learning, in: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

- [83] J. Zhang, J. Pu, C. Chen, R. Fleischer, Low-resolution gait recognition, *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 40 (4) (2010) 986–996.
- [84] G. Dai, D. Y. Yeung, Tensor embedding methods, in: *Proc. Twenty-First National Conference on Artificial Intelligence*, 2006, pp. 330–335.
- [85] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (1) (2007) 40–51.
- [86] D. Xu, S. Lin, S. Yan, X. Tang, Rank-one projections with adaptive margins for face recognition, *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics* 37 (5) (2007) 1226–1236.
- [87] X. Gao, X. Li, J. Feng, D. Tao, Shot-based video retrieval with optical flow tensor and HMMs, *Pattern Recognition Letters* 30 (2) (2010) 140–147.
- [88] S. Aja-Fernández, R. d. L. García, D. Tao, X. Li (Eds.), *Tensors in Image Processing and Computer Vision*, Springer, 2009.
- [89] A. Cichocki, R. Zdunek, A. H. Phan, S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley-Blackwell, 2009.
- [90] Workshop on Algorithms for Modern Massive Data Sets (2010, 2008, 2006). URL <http://www.stanford.edu/group/mmds/>
- [91] A. K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video Technology* 14 (1) (2004) 4–20.
- [92] J. Wang, A. Barreto, L. Wang, Y. Chen, N. Rishe, J. Andrian, M. Adjouadi, Multilinear principal component analysis for face recognition with fewer features, *Neurocomputing* 73 (10-12) (2010) 1550–1555.
- [93] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Uncorrelated multilinear principal component analysis through successive variance maximization, in: *Proc. Int. Conf. on Machine Learning*, 2008, pp. 616–623.
- [94] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Uncorrelated multilinear discriminant analysis with regularization for gait recognition, in: *Proc. Biometrics Symposium 2007*, 2007, doi:10.1109/BCC.2007.4430540.
- [95] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Boosting LDA with regularization on MPCA features for gait recognition, in: *Proc. Biometrics Symposium 2007*, 2007, doi:10.1109/BCC.2007.4430542.
- [96] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Multilinear principal component analysis of tensor objects for recognition, in: *Proc. Int. Conf. on Pattern Recognition*, Vol. 2, 2006, pp. 776 – 779.

- [97] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, Boosting discriminant learners for gait recognition using mpca features, *EURASIP Journal on Image and Video Processing* 2009, article ID 713183, 11 pages, doi:10.1155/2009/713183.
- [98] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Processing Magazine* 25 (1) (2008) 41–56.
- [99] J. Wen, X. Gao, Y. Yuan, D. Tao, J. Li, Incremental tensor biased discriminant analysis: A new color-based visual tracking method, *Neurocomputing* 73 (4-6) (2010) 827–839.
- [100] H. Lu, H.-L. Eng, M. Thida, K. N. Plataniotis, Visualization and clustering of crowd video content in mpca subspace, in: *Proc. 19th ACM Conference on Information and Knowledge Management*, 2010, pp. 1777–1780.
- [101] X. Gao, Y. Yang, D. Tao, X. Li, Discriminative optical flow tensor for video semantic analysis, *Computer Vision and Image Understanding* 113 (3) (2009) 372–383.
- [102] J. Wen, X. Li, X. Gao, D. Tao, Incremental learning of weighted tensor subspace for visual tracking, in: *Proc. 2009 IEEE Int. Conf. on Systems, Man and Cybernetics*, 2009, pp. 3688–3693.
- [103] J. A. Ruiz-Hernandez, J. L. Crowley, A. Lux, “How old are you?”: Age estimation with tensors of binary gaussian receptive maps, in: *Proceedings of the British Machine Vision Conference*, 2010, pp. 6.1–11.
- [104] J. B. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 290 (22) (2000) 2319–2323.
- [105] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (22) (2000) 2323–2326.
- [106] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (3) (2005) 328–340.
- [107] D. Cai, X. He, J. Han, H. J. Zhang, Orthogonal Laplacianfaces for face recognition, *IEEE Transactions on Image Processing* 15 (11) (2006) 3608–3614.
- [108] K. W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, *Computer Vision and Image Understanding* 101 (1) (2006) 1–15.
- [109] S. Z. Li, C. Zhao, M. Ao, Z. Lei, Learning to fuse 3D+2D based face recognition at both feature and decision levels, in: *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, 2005, pp. 43–53.
- [110] A. Colombo, C. Cusano, R. Schettini, 3D face detection using curvature analysis, *Pattern Recognition* 39 (3) (2006) 444–455.

- [111] P. J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, 2005, pp. 947–954.
- [112] R. Xu, D. W. II, Survey of clustering algorithms, IEEE Transactions on Neural Networks 16 (3) (2005) 645–678.
- [113] C. Nolker, H. Ritter, Visual recognition of continuous hand postures, IEEE Transactions on Neural Networks 13 (4) (2002) 983–994.