Neurocomputing 129 (2014) 17–24

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multifactor sparse feature extraction using Convolutive Nonnegative Tucker Decomposition

Qiang Wu [a,*], Liqing Zhang [b], Andrzej Cichocki [c]

[a] School of Information Science and Engineering, Shandong University, Jinan, Shandong, China
[b] MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
[c] Laboratory for Advanced Brain Signal Processing, BSI RIKEN, Wakoshi, Saitama, Japan and Warsaw University of Technology Department of EE, Poland

ARTICLE INFO

Article history:
Received 31 January 2012
Received in revised form
17 December 2012
Accepted 6 April 2013
Available online 24 October 2013

Keywords:
Nonnegative Tensor Decomposition
Convolutive Tucker Model
Alternating Least Squares (ALS)
Feature extraction
Robustness

ABSTRACT

Multilinear algebra of the higher-order tensor has been proposed as a potential mathematical framework for machine learning to investigate the relationships among multiple factors underlying the observations. One popular model Nonnegative Tucker Decomposition (NTD) allows us to explore the interactions of different factors with nonnegative constraints. In order to reduce degeneracy problem of tensor decomposition caused by component delays, convolutive tensor decomposition model is an appropriate model for exploring temporal correlations. In this paper, a flexible two stage algorithm for *K*-mode Convolutive Nonnegative Tucker Decomposition (*K*-CNTD) model is proposed using an alternating least square procedure. This model can be seen as a convolutive extension of Nonnegative Tucker Decomposition. The patterns across columns in convolutive tensor model are investigated to represent audio and image considering multiple factors. We employ the *K*-CNTD algorithm to extract the shift-invariant sparse features in different subspaces for robust speaker recognition and Alzheimer's Disease(AD) diagnosis task. The experimental results confirm the validity of our proposed algorithm and indicate that it is able to improve the speaker recognition performance especially in noisy conditions and has potential application on AD diagnosis.

1. Introduction

Multilinear algebra provides a powerful data modeling framework for exploring data with multiple factors. It has a wide applications, ranging from machine learning to signal processing and beyond [1–6]. Widely used tensor decomposition methods include PARAFAC model [7], Tucker model [8], Nonnegative Tensor Factorization [1] which imposes the nonnegative constraint on the PARAFAC or Tucker model. Furthermore, extended tensor decomposition models INDSCAL, DEDICOM [1,9,10] are proposed to explore symmetry in tensors and Block Term Decomposition and CANDELING considering models interpolating between PARAFAC and Tucker models. Compared with traditional matrix factorization methods, tensor decomposition models are suitable to preserve the natural structures of higher order data.

Several widely used algorithms based on Tucker model have imposed orthogonal constraints on factors for the feature extraction or data mining tasks. For example, De Lathauwer [11] proposed the Higher-Order Singular Value Decomposition (HOSVD) for tensor decomposition, which is a multilinear generalization of the matrix SVD. Higher Order Orthogonal Iteration (HOOI) [12–14] extended the truncated SVD algorithms to the tensor-structure data. Panagakis [15] developed a new tensor factorization method called Nonnegative Multilinear Principal Components Analysis (NMPCA) to find a tensor-to-tensor projection [16] via multilinear subspace learning for music genre classification. Nonnegative Tucker Decomposition (NTD) [17] is a natural extension of NMF algorithms. The multiplicative algorithm for NTD is based on minimization of the squared Euclidean distance and the KL divergence. Some generalized cost functions based on Alpha-, Beta- and Bregman-divergence [1,18,19] were also used. With regard to optimization solutions, ALS and HALS algorithms [20] were derived from Newton methods and they achieved good convergence rate.

Recently, the degeneracy problem of tensor decomposition [21,22] has been investigated due to the component delays under multiple factor observations. As stated in [22–24], we observe often component delays in many applications based on tensor structure, such as time shifts in fMRI data due to hemodynamic delay, delays across trials in EEG data when onset changes were not locked to the event. The shifted or convolutive tensor decomposition model can been seen as an extension of original model.

* Corresponding author. Tel.: +86 531 88362526.
E-mail addresses: wuqiang@sdu.edu.cn (Q. Wu), lqzhang@sjtu.edu.cn (L. Zhang), cia@brain.riken.jp (A. Cichocki).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.04.049

PARAFAC2 model [25] was proposed to handle retention time shifts in resolving chromatographic data. As an extension of shifted factor analysis, the N-way shifted factor analysis model is investigated in [21,22,26,27]. In [28], Mørup proposed a 2D Convolutive NTF (CNTF) algorithm for multichannel time–frequency analysis. Shift Invariant Sparse Coding (SISC) model [29] is an extension of sparse coding to handle data from linear mixtures. Makkiabadi [30] proposed a generalization of PARAFAC2 model [25] for convolutive mixture. Therefore, there exist a large number of demands on efficient and fast algorithms for shifted or convolutive tensor decomposition model to suit with the practical data better.

In order to reduce the effect of degeneracy problem caused by component delays, we propose a novel K-mode Convolutive Nonnegative Tucker Decomposition (K-CNTD) model as an extension of NTD. A two stage algorithm is developed to estimate the shifted factor matrices and core tensor. In the first stage we employ NTD to factorize the convolutive mixture in tensor structure into factor matrices and core tensor. For the purpose of considering components delays, in the second stage the original components in K modes are recovered by the convolutive NMF algorithms. The efficiency of K-CNTD algorithm is verified on synthetic data, noisy speech signal and AD sMRI data. Extensive simulation results demonstrate that the shift-invariant sparse features extracted by our proposed algorithm are robust for speaker recognition in noisy conditions and efficient to improve the diagnosis/classification performance for Alzheimer's Disease.

The remainder of this paper is organized as follows. In Section 2, the background knowledge about convolutive nonnegative matrix factorization and tensor analysis is introduced. In Section 3, a two stage algorithm for Convolutive Nonnegative Tucker Decomposition model is presented for feature extraction. Section 4 describes the experimental results of synthetic data, robust speaker recognition in noisy environments and AD sMRI diagnosis task. Finally, Section 5 provides a summary and conclusions.

2. Background

2.1. Convolutive nonnegative matrix factorization

Convolutive Nonnegative Matrix Factorization (CNMF) [31] is generalization of NMF by considering the relative position of basis functions or coefficients in feature space. It aims at extracting cross-column patterns as single basis function. First, we introduce the following operations, upward, downward, left and right shifted operators $\overset{k\uparrow}{(A)}, \overset{k\downarrow}{(A)}, \overset{k\leftarrow}{(A)}, \overset{k\rightarrow}{(A)}$ on the matrix A by shifting and zero padding the rows or columns of A. For example

$$\overset{0\rightarrow}{A}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\quad \overset{0\leftarrow}{A}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\quad \overset{1\rightarrow}{A}=\begin{pmatrix}0&1&2\\0&4&5\\0&7&8\end{pmatrix}$$
$$\overset{1\leftarrow}{A}=\begin{pmatrix}2&3&0\\5&6&0\\8&9&0\end{pmatrix}\quad \overset{1\uparrow}{A}=\begin{pmatrix}4&5&6\\7&8&9\\0&0&0\end{pmatrix}\quad \overset{1\downarrow}{A}=\begin{pmatrix}0&0&0\\1&2&3\\4&5&6\end{pmatrix} \tag{1}$$

where the matrix A is

$$A=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}$$

Based on the definition of shifted operators, the CNMF model is defined as

$$V\approx\sum_{l=0}^{L-1}W_l\overset{l\rightarrow}{H} \tag{2}$$

where $V\in\mathbb{R}^{M\times N}\geq 0$ is the input matrix, $W_l|_{l=0}^{L-1}\in\mathbb{R}^{M\times R}\geq 0$ is a set of basis functions and $H\in\mathbb{R}^{R\times N}$ is the weight coefficients.

Model (2) can be decomposed into a set of NMF approximations [31]. The Alternating Least Square(ALS) method has been widely applied to find the decomposed factors. The algorithm is a Newton-like method and has good convergence rate [1]. As described in [32], the ALS update rules of each NMF approximation for W_l and $\overset{l\rightarrow}{H}$ can be derived as

$$H_l\leftarrow[(W_l^TW_l)^{-1}(W_l^T\overset{l\leftarrow}{V})]_+,\quad W_l\leftarrow[\overset{l\rightarrow}{V}\overset{l\rightarrow}{H}^T(\overset{l\rightarrow}{H}\overset{l\rightarrow}{H}^T)^{-1}]_+ \tag{3}$$

where $(\cdot)^T$ is the transpose operator, $[a]_+=\max(\varepsilon,a)$ is a half-wave rectifying nonlinear projection to enforce nonnegativity [32]. For each l, H_l corresponds to $\overset{l\rightarrow}{H}$. The basis function W_l and coefficient matrix H_l are updated for each l. As stated in [31], the algorithm first update all W_l and then final H is assigned to the average of $H_l|_{l=0}^{L-1}$, i.e.

$$H\leftarrow\frac{1}{L}\sum_{l=0}^{L-1}H_l \tag{4}$$

We use the relative error e_{nmf} defined in (5) as a stop criterion of the algorithm:

$$e_{nmf}=\left\|V-\sum_{l=0}^{L-1}W_l\overset{l\rightarrow}{H}\right\|_F/\|V\|_F \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm.

2.2. Multilinear algebra

Multilinear algebra is the algebra of higher order tensors. A tensor is a higher order generalization of matrix. Let $\underline{X}\in\mathbb{R}^{I_1\times I_2\cdots\times I_N}$ denote a tensor. The order of $\underline{X}$ is N. The mode-n matricization of an N order tensor X rearranges the elements of $\underline{X}$ to form the matrix $X_{(n)}\in\mathbb{R}^{I_n\times I_{n+1}\overline{I_{n+2}\cdots I_N I_1\cdots I_{n-1}}}$.

The n-mode product of a tensor $\underline{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ and matrix $A\in\mathbb{R}^{J_n\times I_n}$ is denoted by $\underline{Y}=\underline{X}\times_nA\in\mathbb{R}^{I_1\times\cdots\times I_{n-1}\times J_n\times I_{n+1}\cdots\times I_N}$ and it is defined as

$$y_{i_1,i_2,\ldots i_{n-1}j_n,i_{n+1},\ldots i_N}=\sum_{i_n}x_{i_1,\ldots i_n,\ldots,i_N}a_{j_n,i_n} \tag{6}$$

In this paper we simplify the notation as

$$\underline{G}\times_1A^{(1)}\times\cdots\times_NA^{(n)}=\underline{G}\prod_{n=1}^{N}\times_nA^{(n)} \tag{7}$$

The Frobenius norm of a tensor $\underline{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ [33] is given by

$$\|\underline{X}\|_F=\sqrt{\sum_{i_1=1}^{I_1}\cdots\sum_{i_N=1}^{I_N}x_{i_1,\ldots,i_N}^2} \tag{8}$$

Obviously the mode-n matricization of tensor $X_{(n)}$ has the same Frobenius norm as tensor $\underline{X}$ that is $\|X_{(n)}\|_F=\|\underline{X}\|_F$.

Some basic notations of multilinear algebra are described in Table 1. The details about tensor decomposition can be found in [1,11,32,33].

2.3. Nonnegative Tucker Decomposition

Nonnegative Tucker Decomposition (NTD) [17] model is defined as

$$\underline{X}=\underline{G}\times_1U^{(1)}\times_2U^{(2)}\cdots\times_NU^{(N)}+\underline{E} \tag{9}$$

Table 1
Notations in multilinear algebra.

Notation	Description
$\otimes$	Kronecker product
$\circledast$	Hadamard product
$\oslash$	Element-wise division
X	Matrix
$U^{(n)}$	The nth factor after tensor factorization
$\underline{X}$	Tensor
$X_{(n)}$	n-mode matricized of tensor $\underline{X}$
$\times_n$	n-mode product of tensor and matrix
$U^{\otimes}$	$U^{(N)} \otimes \cdots \otimes U^{(1)}$
$U^{\otimes_{-n}}$	$U^{(N)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)}$

where $\underline{X} \in \mathbb{R}_+^{I_1 \times I_2 \cdots \times I_N} \geq 0$ is the data tensor, $\underline{G} \in \mathbb{R}_+^{J_1 \times \cdots \times J_N} \geq 0$ is the core tensor, $U^{(n)}|_{n=1}^{N} \in \mathbb{R}_+^{I_n \times J_n} \geq 0$ is a set of nonnegative factor matrices, $\underline{E}$ is the residual tensor. Equivalently, NTD model can be written in matrix notation by use of Kronecker product as

$$X_{(n)} = U^{(n)} G_{(n)} U^{\otimes_{-n}T} + E_{(n)} \tag{10}$$

As described in [1,20,32], the ALS update rules for factor matrices $U^{(n)}|_{n=1}^{N}$ and core tensor $\underline{G}$ are given by

$$U^{(n)} \leftarrow [X_{(n)} U^{\otimes_{-n}} G_{(n)}^T (G_{(n)} (U^T U)^{\otimes_{-n}} G_{(n)}^T)]_+ \tag{11}$$

$$\underline{G} \leftarrow \underline{G} \circledast \left[\left(\underline{X} \prod_{n=1}^{N} \times_n U^{(n)T} \right) \oslash \left(\underline{G} \prod_{n=1}^{N} \times_n U^{(n)T} U^{(n)} \right) \right] \tag{12}$$

3. Two stage algorithm for *K*-mode Convolutive Nonnegative Tucker Decomposition

The component shifts or delays have been considered in many areas of science [21]. For example, audio signal in the reverberant environment exists the temporal shifts that cause the cocktail party problem. Shifting of absorption and emission spectra occur in chemistry and physics. There also exist component delays in fMRI and EEG data. So current tensor decomposition model without considering shifting will cause the model mismatch with the data. As stated in [21,22,34], the degeneracy problem of tensor decomposition model will occur due to the component delays.

In order to consider the potential dependencies across the columns of factor matrices and investigate component delay patterns that span multiple columns of factor matrices, we extend NTD model into convolutive form. Considering the delays in first mode, we write the convolutive NTD model in one mode as

$$\underline{X} = \sum_{l=0}^{L-1} \underline{G}_l \times_1 \overset{l\downarrow}{H^{(1)}} \times_2 U^{(2)} \times \cdots \times_N U^{(N)} + \underline{E} \tag{13}$$

where the original data tensor $\underline{X}$ is decomposed into a set of core tensor $\underline{G}_l|_{l=0}^{L-1}$, factor matrices $U^{(n)}|_{n=2}^{N}$ and shifted factor matrix $H^{(1)}$. The core tensors can be seen as a set of higher order basis functions. $U^{(n)}|_{n=2}^{N}$ and $H^{(1)}$ are the weights or coefficients. Especially, the basis functions with higher order tensor structure will be shifted and scaled in the first mode by convolution across the axis of l with the rows of $H^{(1)}$. Our objective is to estimate the appropriate set of core tensors $\underline{G}_l|_{l=0}^{L-1}$, factor matrices $U^{(n)}|_{n=2}^{N}$ and $H^{(1)}$ to approximate data tensor $\underline{X}$. In order to simplify the estimation process, we can decompose the set of core tensors into a intermediate common tensor $\underline{G}$ and a set of matrices $W_l^{(1)}|_{l=0}^{L-1}$, i.e. $\underline{G}_l = \underline{G} \times_1 W_l^{(1)}, l = 0, \ldots, L-1$. Then we can obtain the equivalent expression of Eq. (13) as follows:

$$\begin{aligned}
\underline{X} &= \sum_{l=0}^{L-1} (\underline{G} \times_1 W_l^{(1)}) \times_1 \overset{l\downarrow}{H^{(1)}} \times_2 U^{(2)} \times \cdots \times_N U^{(N)} + \underline{E} \\
&= \underline{G} \times_1 \sum_{l=0}^{L-1} \left(\overset{l\downarrow}{H^{(1)}} W_l^{(1)} \right) \times_2 U^{(2)} \times \cdots \times_N U^{(N)} + \underline{E} \\
&= \underline{G} \times_1 U^{(1)} \times_2 U^{(2)} \times \cdots \times_N U^{(N)} + \underline{E}
\end{aligned} \tag{14}$$

where $U^{(1)} = \sum_{l=0}^{L-1} \overset{l\downarrow}{H^{(1)}} W_l^{(1)}$, the common core tensor $\underline{G}$ is the multiple factor basis functions of tensor $\underline{X}$ without considering component delays. From Eq. (14), we find the estimation procedure can be separated into two stages. In the first stage, we can use NTD algorithm to estimate the common core tensor $\underline{G}$ and factor matrices $U^{(n)}|_{n=1}^{N}$. In the second stage, the intermediate matrix $U^{(1)}$ can be decomposed into $H^{(1)}$ and $W_l^{(1)}|_{l=0}^{L-1}$ by convolutive NMF algorithm. Then the higher order basis functions $\underline{G}_l|_{l=0}^{L-1}$ can be estimated by $\underline{G}_l = \underline{G} \times_1 W_l^{(1)}, l = 0, \ldots, L-1$. Fig. 1 illustrates the model of convolutive NTD in one mode.

More generally, we can extend Eq. (13) into convolutive form in K modes as

$$\underline{X} = \sum_{l_1=0}^{L_1-1} \cdots \sum_{l_K=0}^{L_K-1} \underline{G}_{l_1 \ldots l_K} \times_1 \overset{l_1\downarrow}{H^{(1)}} \times \cdots \times_K \overset{l_K\downarrow}{H^{(K)}} \times_{K+1} U^{(K+1)} \times \cdots \times_N U^{(N)} + \underline{E} \tag{15}$$

where $\underline{G}_{l_1 \ldots l_K}$, $l_k = 0, 1, \ldots, L_k - 1$, $k = 1, \ldots, K$ are the higher order basis functions, $H^{(k)}|_{k=1}^{K}$ are the shifted factor matrices and $U^{(n)}|_{n=K+1}^{N}$ are the factor matrices. Similar to Eq. (14), we can derive the following expression:

$$\begin{aligned}
\underline{X} &= \sum_{l_1=0}^{L_1-1} \cdots \sum_{l_K=0}^{L_K-1} \underline{G}_{l_1 \cdots l_K} \times_1 \overset{l_1\downarrow}{H^{(1)}} \times \cdots \times_K \overset{l_K\downarrow}{H^{(K)}} \times_{K+1} U^{(K+1)} \times \cdots \times_N U^{(N)} + \underline{E} \\
&= \sum_{l_1=0}^{L_1-1} \cdots \sum_{l_K=0}^{L_K-1} \left(\underline{G} \prod_{k=1}^{K} \times_k W_{l_k}^{(k)} \right) \times_1 \overset{l_1\downarrow}{H^{(1)}} \times \cdots \times_K \overset{l_K\downarrow}{H^{(K)}} \times_{K+1} U^{(K+1)} \\
&\quad \times \cdots \times_N U^{(N)} + \underline{E} \\
&= \underline{G} \times_1 \left(\sum_{l_1=0}^{L_1-1} \overset{l_1\downarrow}{H^{(1)}} W_{l_1}^{(1)} \right) \times \cdots \times_K \left(\sum_{l_K=0}^{L_K-1} \overset{l_K\downarrow}{H^{(K)}} W_{l_K}^{(K)} \right) \times_{K+1} U^{(K+1)} \\
&\quad \times \cdots \times_N U^{(N)} + \underline{E} = \underline{G} \times_1 U^{(1)} \times_2 U^{(2)} \times \cdots \times_N U^{(N)} + \underline{E}
\end{aligned} \tag{16}$$

where the core tensors $\underline{G}_{l_1 \ldots l_K} = \underline{G} \prod_{k=1}^{K} \times_k W_{l_k}^{(k)}$, $l_k = 0, 1, \ldots, L_k - 1$, $k = 1, \ldots, K$, the intermediate matrices $U^{(k)} = \sum_{l_1=0}^{L_1-1} \overset{l_1\downarrow}{H^{(1)}} W_{l_1}^{(1)}$, $k = 1, \ldots, K$. According to Eq. (16), we can derive the two stage algorithm for K-Mode Convolutive NTD (K-CNTD) based on the ALS NTD [1,20] and ALS convolutive NMF algorithms. The algorithm in detail is described in Algorithm 1.

Algorithm 1. Algorithm for K-CNTD.

Input:

Given data tensor $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N} \geq 0$, the components number $\{J_n\}_{n=1}^{N}$ for NTD, the convolutive length L_k, the components number T_k for CNMF, $(k = 1, \ldots, K)$.

Output:

The estimated components $W_{l_k}^{(k)}|_{l_k=0,k=1}^{L_k,K}$, $H^{(k)}|_{k=1}^{K}$, $U^{(n)}|_{n=K+1}^{N}$, $\underline{G}_{l_1, \ldots, l_K}$.

1: Initialization: Set $U_0^{(n)}|_{n=1}^{N}, \underline{G}_0$ randomly, normalize all $U_0^{(n)}|_{n=1}^{N}$;
2: **repeat**
3: **for** $n = 1 : N$ **do**
4: % Update $U^{(n)}$
5: $U^{(n)} \leftarrow [X_{(n)} U^{\otimes_{-n}} G_{(n)}^T (G_{(n)} (U^T U)^{\otimes_{-n}} G_{(n)}^T)]_+$;

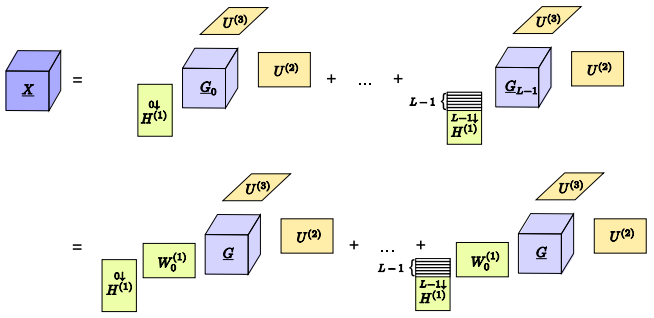

Fig. 1. Convolutive NTD model in one mode.

6: **end for**
7: % Update core tensor $\underline{G}$
8: $\underline{G} \leftarrow \underline{G} \circledast \left[\left(\underline{X} \prod_{n=1}^{N} \times_n U^{(n)T} \right) \oslash \left(\underline{G} \prod_{n=1}^{N} \times_n U^{(n)T} U^{(n)} \right) \right]$;
9: **until** $\|\underline{X} - \underline{G} \prod_{n=1}^{N} \times_n U^{(n)}\|_F / \|\underline{X}\|_F < \varepsilon$
10: **for** $k=1:K$ **do**
11: Set $W_{l_k}^{(k)}|_{l_k=0}^{L_k-1}$ and $H^{(k)}$ randomly;
12: **repeat**
13: **for** $l_k=0:L_k-1$ **do**
14: $$H_{l_k}^{(k)} \leftarrow \left[\left(W_{l_k}^{(k)T} W_{l_k}^{(k)} \right)^{-1} \left(W_{l_k}^{(k)T} U^{(k)T} \overleftarrow{{}^{l_k}} \right) \right]_+ ;$$
15: $$W_{l_k}^{(k)} \leftarrow \left[U^{(k)T} \left(H^{(k)} \overrightarrow{{}^{l_k}} \right)^T \left(H^{(k)} \left(H^{(k)} \overrightarrow{{}^{l_k}} \right)^T \right)^{-1} \right]_+ ;$$
16: **end for**
17: $H^{(k)} \leftarrow \frac{1}{L} \sum_{l_k=0}^{L_k-1} H_{l_k}^{(k)}$;
18: **until** $\|U^{(k)T} - \sum_{l_k=0}^{L_k-1} W_{l_k}^{(k)} H^{(k)} \overrightarrow{{}^{l_k}} \|_F / \|U^{(k)T}\|_F < \varepsilon$
19: **end for**
20: $\underline{G}_{l_1,\ldots,l_K} = \sum_{l_1} \cdots \sum_{l_k} \underline{G} \times_1 W_{l_1}^{(1)} \cdots \times_N W_{l_K}^{(K)}$;

Here we derive that the alternating least square algorithm for K-CNTD model and the update rules with half-wave rectifying nonlinear projection for ALS CNMF and NTD are similar to the Exponential Gradient in NMF. The monotonic convergence analysis in [35,36] can be applied to our case as well. And from the experimental result, our proposed algorithm also shows good convergence as proved in [36].

The proposed K-CNTD algorithm is able to extract the repeating patterns across columns in given modes. When analyzing audio or image data, such dependencies across successive columns are frequently explored. Here, we investigate the repeating patterns in multifactor form as a new desire feature for audio or image representation. These features have the expressive ability to capture the temporal or frequency dependencies within a set of convolutive higher order basis functions.

4. Simulation

In this section, we present the simulation results on synthetic data, robust speaker recognition and AD diagnosis task using K-CNTD algorithm. The proposed algorithm is effective for complex feature extraction task by identifying hidden components.

4.1. Synthetic data

In order to evaluate the K-CNTD algorithm in term of its effectiveness, a simulation study on synthetic data was undertaken.

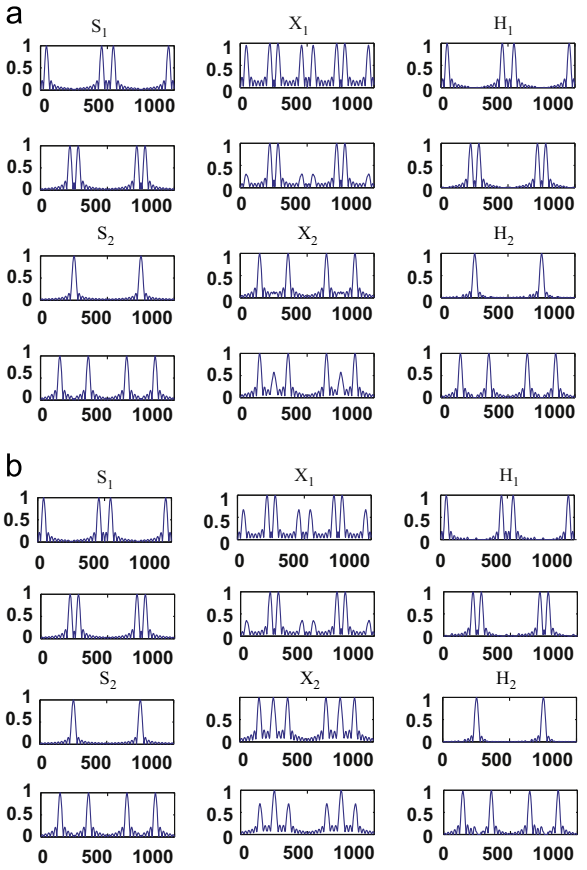

Fig. 2. Estimated results with convolution length $L=2$ and 4.

We used $S_1 \in \mathbb{R}^{2\times 1000}$ and $S_2 \in \mathbb{R}^{2\times 1000}$ as sources signal to generate convolutive mixture X_1 and X_2 respectively. Several samples of S_1 and S_2 are shown in Fig. 2. The convolutive mixture $X_k = \sum_{l_k=0}^{L_k-1} A_{l_k}^k \overrightarrow{S}^{l_k}{}_k$, $k=1,2$, where $A_{l_k}^k|_{l_k=0}^{L_k-1}$ are the mixture matrices.

We used X_1, X_2, $X_3 \in \mathbb{R}^{2\times 2}$ and $\underline{G} \in \mathbb{R}^{2\times 2\times 2}$ to generate a 3-order tensor $\underline{X}_{test} \in \mathbb{R}^{1000\times 1000\times 2}$ which can be seen as a mixture procedure in tensor structure by factor matrix X_3 and core tensor $\underline{G}$, i.e.

$$\underline{X}_{test} = \underline{G} \times_1 X_1 \times_2 X_2 \times_3 X_3 \tag{17}$$

We employed K-CNTD to recover the sources components and the estimated components were denoted as $H_k|_{k=1}^2$. Fig. 2 gives the estimated signal with the convolution length $L=2$ and 4. From this result, K-CNTD algorithm can recover the original signal from the tensor mixture.

4.2. Speaker recognition in noisy conditions

In this experiment we applied K-CNTD algorithm to extract robust features for the speaker recognition task in noisy conditions. Grid corpus (speech of 34 persons) mixed with different noise was used to test the recognition performance. We employed the cortical-based feature extraction framework described in [37] with 4-order tensor structure (time × frequency × scale × direction) and K-CNTD algorithm to extract the shift-invariant sparse speech features in time–frequency domain. We employed following steps to extract the robust speech features:

1. Suppose that the speech signal is denoted by $s(t)$, we first perform pre-emphasis and Short Time Fourier Transformation

(STFT) on the speech signal and calculate the power spectrum $S(f,t)$.
2. Employ the Gabor filtering with different scales and directions and Mel filtering to filter $S(f,t)$ and obtain the cortical representation $\underline{S}$.
3. Calculate the shifted factor matrices in different modes and core tensor set using *K*-CNTD algorithm; using factor matrix $H^{(2)}$ in frequency mode we projected the cortical representation into feature subspace and calculate the sparse tensor features $\underline{Y}$.
4. Unfold tensor $\underline{Y}$ into feature matrix S_R and employ Discrete Cosine Transform (DCT) to reduce the dimension.

The sampling rate of speech signal was 8 kHz. A hamming window of 25 ms was shifted over an input speech utterance every 10 ms to calculate power spectrum. At each window position, a segmented utterance was converted to its corresponding 256-dimensional FFT-based power spectrum vector. Gabor filters with four different scales and four different directions were employed to derive the multiresolution Gabor-based features from power spectrum. Then the multifactor Gabor features were filtered by 40-channel Mel filterbanks to create the cortical representation for tensor decomposition. *K*-CNTD algorithm was employed to decompose the Gabor-based tensor data to obtain the shift factor matrix. The component number $\{J_n\}_{n=1}^4$ for NTD were 50, 25, 3 and 3 and component number T_1 for time mode and T_2 for frequency mode were all 20. The convolutive length in time and frequency modes were all set to 3.

We randomly selected 1700 sentences (50 sentences for each speaker) as training data and testing data includes 10 sets, each set contained 2040 sentences (60 sentences for each speaker). The testing samples in noisy conditions were generated by mixing with Babble, Destroyer engine, Buccaneer, Factory, Pink, White noises in SNR intensities of -5 dB, 0 dB, 5 dB and 10 dB respectively. The basis function $H^{(2)}$ in frequency mode was used to project the cortical representation into feature subspace and obtain the sparse tensor features. The final feature vectors were extracted by DCT with 16 cepstral coefficients. GMM with 64 Gaussian mixtures was employed as the recognizer for speaker modeling.

For comparison, we tested the performance of MFCC, PLP, Spectral Substraction (SS), CNMF and NTD. For MFCC and PLP, the windows width and overlap length was the same as *K*-CNTD algorithm-based method. After 40-channel Mel filterbanks filtering and DCT, MFCC features were obtained. PLP features with 8-order model were calculated by power spectrum after RASTA filtering and DCT. The speech enhancement method spectral substraction proposed in [38] was used to reduce the noise component with initial silence 0.25 s. CNMF algorithm was employed to extract the spectral-temporal features after fixed scale and directions Gabor filtering from power spectrum. NTD-based feature extraction procedure was the same as our proposed framework. The component number $\{J_n\}_{n=1}^4$ were 50, 25, 3 and 3.

Fig. 3 gives the DCT feature comparison between MFCC and features extracted by *K*-CNTD in clean and 5 dB conditions. The degradation of MFCC is evident. Compared with the clean condition, the shift-invariant features extracted by *K*-CNTD maintain the useful information and provide robust and natural representation for speaker modeling.

We summarize the average recognition accuracy of *K*-CNTD and baseline systems in all conditions in Fig. 4. The speaker recognition performance using *K*-CNTD is tested on six different noises with various SNR (-5, 0, 5 and 10 dB). Final recognition accuracy in each SNR with different noises is averaged on 10 different testing sets. The accuracy in six noisy conditions averaged over SNRs between -5 and 10 dB, and the overall average accuracy across all the conditions is presented in Fig. 4. These results suggest that our proposed *K*-CNTD algorithm can give a better average recognition result than NTD algorithm and traditional feature extraction methods.

It is observed that the features extracted by *K*-CNTD perform significantly better in the presence of white and destroyer engine noise and slightly better in the presence of babble and factory noise. The speech signal mixed with babble noise consists of other humans' speech signals. The noisy components corrupt the entire frequency bands and also share the statistical properties of the reference signal. So the performance of our proposed method in babble noise condition degrades compared with other noises such as white noise, although the recognition accuracy of our method is still better than the baseline methods. For the other types of noise sources such as white and destroyer engine, their statistical characteristics that *K*-CNTD algorithm utilizes to extract robust features are quite different from that of reference statistics.

4.3. Diagnosis of Alzheimer's disease by sMRI with convolutive tensor model

Alzheimer's disease is the most common cause of dementia that leads to progressive loss of memory and cognition function. Its early and accurate diagnosis/classification is important for the disease prevention. In this experiment, we applied *K*-CNTD algorithm to analyze structural magnetic resonance imaging (sMRI) data of AD subjects and Health Control (HC) subjects. Efficient features for classification of AD and HC were extracted. The performance of classification was tested on the freely public brain imaging data from OASIS [39]. Two groups subjects were selected: 100 AD subjects (the CDR score greater than 0, 59 females and 41 males) and 109 HC subjects (62 females and 47 males).

For two groups sMRI data, we realigned all images to the first image. Then, the sMRI images of all subjects were normalized into a standard space defined by T1 template image provided by SPM8 toolbox. After normalization, the sMRIs were re-sliced and smoothed into $2\times2\times2$ mm^3 voxel-size images.

Based on these normalized sMRI images, we constructed a 4-order tensor $\underline{X}\in R^{81\times97\times83\times209}$ with four different modes: coordinates *x*,*y*,*z* and *subjects*. Then *K*-CNTD algorithm was employed to decompose tensor $\underline{X}$ to obtain the higher order basis functions and factor matrices. The convolutive lengths of each mode were 5,5,5 and 2 respectively. The component number in subjects mode was 60. We regarded the row of shifted factor matrix $H^{(4)}\in R^{209\times60}$ as feature vector for each subject.

We separated the AD and HC data into training set and testing set respectively. The training set included 90% feature samples and the testing set was the remaining 10% samples. Finally, we built the SVM classifier to distinguish AD and HC subjects. The training and testing procedures were repeated over 100 times by randomly selecting training and testing samples. The classification framework based on convolutive tensor model is shown in Fig. 5.

In order to evaluate the performance of our proposed method, the accuracy, sensitivity, specificity of classification were calculated and the last two were defined as

$$Sensitivity=\frac{TP}{TP+FN},\quad Specificity=\frac{TN}{TN+FP} \tag{18}$$

where TP is the number of true positives (AD subjects classified correctly), TN is the number of true negatives (HC subjects classified correctly), FP is the number of false positives (HC subjects classified as AD subjects), FN is the number of false negatives (AD classified as HC subjects).

For comparison, CNMF and NTD algorithm was applied to test the classification performance as baseline system. The 3D sMRI images of all subjects are vectorized to construct data matrix with

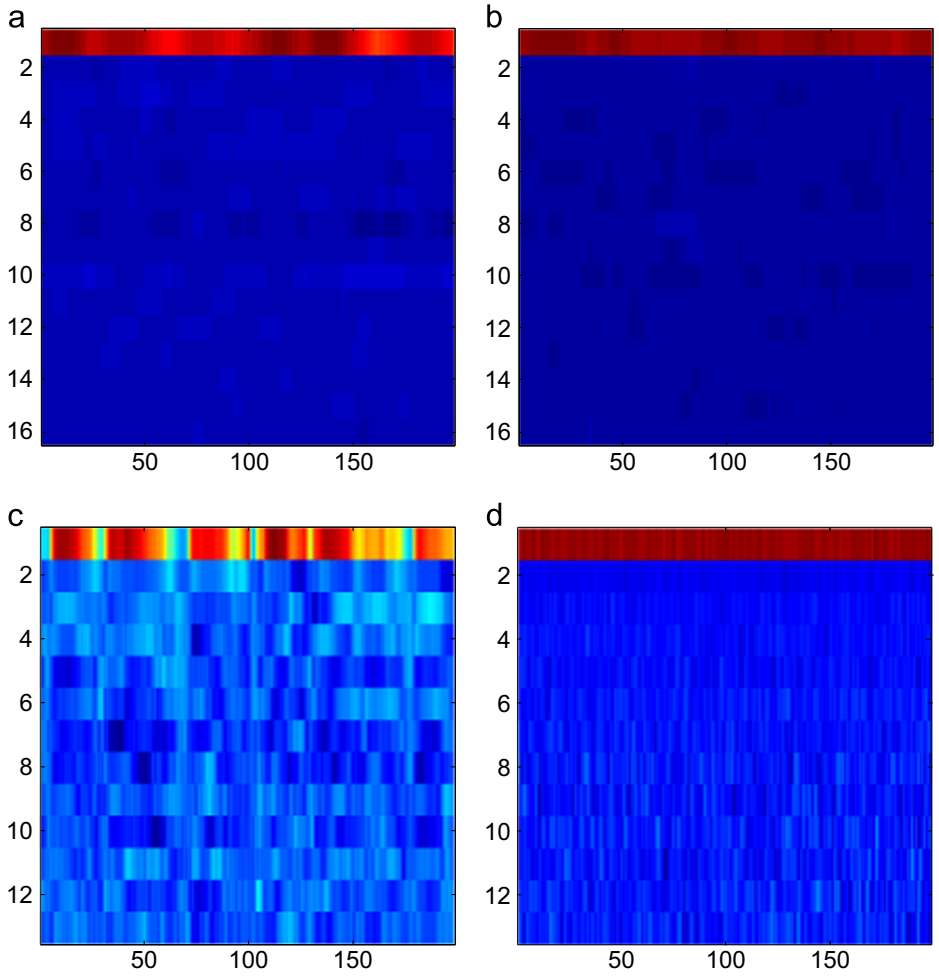

Fig. 3. (a) Clean feature extracted by *K*-CNTD. (b) Feature extracted by *K*-CNTD in 5 dB condition with pink noise. (c) Clean MFCC. (d) MFCC in 5 dB condition with pink noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

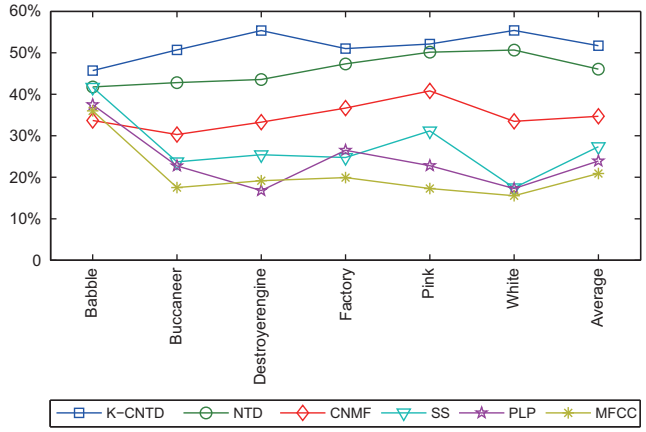

Fig. 4. Average speaker recognition accuracy in different noisy conditions.

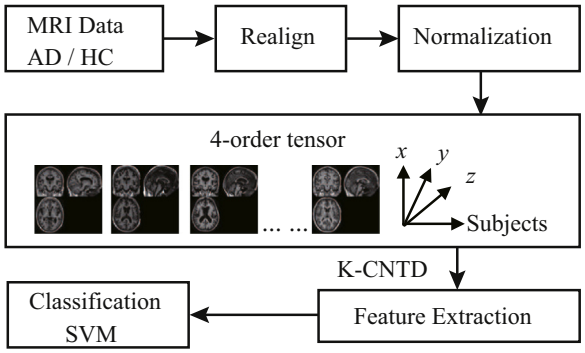

Fig. 5. Classification framework for MRI data.

two dimensions *subjects* and *samples* as input data of CNMF algorithm. We set the convolutive length as 3 and extract the rows of basis functions as feature vectors for training SVM classifier. The feature extraction and classification procedure of NTD algorithm was similar to our proposed framework in Fig. 5. The evaluation results of *K*-CNTD and baseline system, which was the mean of accuracy, sensitivity, specificity for 100 times repeating were summarized in Fig. 6. From the experimental results, classification levels in the range of 80–95% were achieved. Especially, the performance of *K*-CNTD algorithm was over 90% which is better than CNMF and NTD algorithms. This indicated that the shift-invariant sparse feature in multifactor form extracted by *K*-CNTD algorithm was more efficient than CNMF and NTD for distinguishing the AD subjects with HC subjects. It showed that the proposed diganosis/classification framework has big potential for the early AD diagnosis.

4.4. Discussions

In this paper, we present a flexible convolutive tensor decomposition algorithm. Compared with Tucker decomposition model,

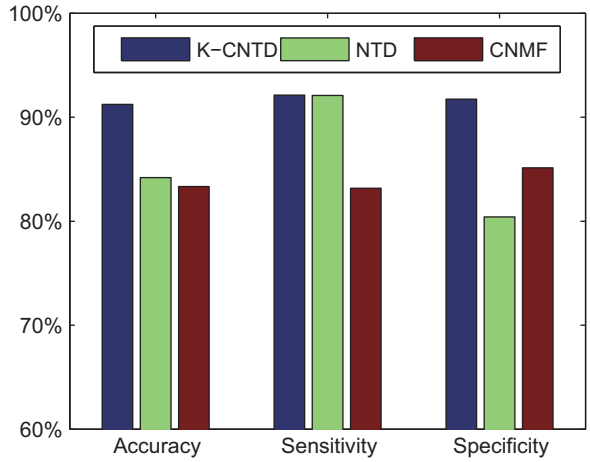

Fig. 6. The average accuracy, sensitivity and specificity results for AD and HC classifications.

our algorithm considers the component delays in given modes, to fit with the practical data better. The extracted features are able to preserve the intrinsic features in the natural structure of data through the multifactor analysis. A two stage algorithm is presented for estimation of *K*-CNTD model. We employ the alternate least square method to estimate desired factors.

When the input data has higher order complex pattern and limited number of samples for training, the linear subspace convolutive model like CNMF will be inadequate to deal with the data in tensor structure. CNMF usually represents the higher order data as vectors or matrices and finds an optimal linear mapping to lower-dimensional space by iteration procedures. The vectorization or matricization of data will destroy the essential structure and correlation in original tensor data. *K*-CNTD algorithm aims to find the optimal decomposition for each factors by keeping the input data in their natural higher order form. Furthermore the experimental results confirm the superiority of *K*-CNTD algorithm compared with CNMF for audio and image feature extraction tasks.

The proposed algorithm discovers qualitatively similar higher order basis functions with NTD algorithm. The difference is that a set of convolutive basis functions is extracted for repeating patterns across columns in given modes. These basis functions encode a lot of information about the speech or image data and naturally reflect particular patterns. For speech signals, these basis functions in frequency mode are representing harmonic series with various inflections and consonant sounds. For images, repeating patterns in these basis functions with sparse constraint recover the truly localized, parts-based components.

Based on the cortical representation, we use 2D Gabor filtering with different scales and directions to simulate the receptive field in cortical simple cells. These representations describe the neuron response for different cues of perceptions. By *K*-CNTD algorithm, the intrinsic features of different factors can be extracted after projection and feature selection.

According to the auditory neural coding [40], we assume that the speech data in the feature space is sparse. Sparse coding theory [41] assumes that given a sound stimulus, only a few auditory neurons are active (nonzero elements) simultaneously. The activity of neurons with small absolute values are regarded as noise and can be set to zero, only a few components with strong activities are considered. The shift-invariance sparse assumption in our proposed method is similar to sparse coding shrinkage method [41]. The sparse assumption can make the feature robust because the energy of clean signal is concentrated on a few components only, while the energy of noises spreads on all the components. From the experimental results, the features extracted by *K*-CNTD algorithm provide better average performance than CNMF and NTD algorithms and traditional feature extraction methods. This result indicates that the *K*-CNTD algorithm can extract more robust shift-invariance sparse features for speaker recognition in noise conditions.

We model the sMRI data of AD and HC subjects as 4-order tensor with four modes (*x*,*y*,*z* and *subjects*). The final feature set is the coefficients of shifted factor matrix in subjects mode. The simulation results show that *K*-CNTD algorithm provides better diagnosis/classification performance compared with NTD algorithm under same feature extraction framework. This indicates that the shift-invariance sparse features extracted by convolutive model are more distinguishable for AD and HC subjects classification. The proposed method discovers more precisely hidden patterns for sMRI image feature extraction.

5. Conclusions

In this paper, we investigate the component delays model for tensor decomposition. A two stage ALS algorithm for *K*-mode Convolutive Nonnegative Tucker Decomposition model is developed to reduce the degeneracy problem caused by component delays. Our proposed model is an extension of Nonnegative Tucker Decomposition and can preserve the intrinsic information in the natural structure of tensor data. We applied *K*-CNTD algorithm for robust speaker recognition and early AD disease diagnosis task. Based on cortical representation of speech signal, multifactor shift-invariant sparse features were extracted by reduce noisy components and improve the robustness of speaker recognition system. By the convolutive model, *K*-CNTD algorithm extracts more discriminative sparse features for AD and HC subjects classification. The final simulation results demonstrated that our proposed algorithm is more efficient for robust speaker recognition and early AD diagnosis compared with the baseline methods.

Acknowledgment

The authors would like to thank anonymous reviewers for their constructive comments on this paper. The work was supported by the National Natural Science Foundation of China (Grant nos. 61305060, 61272251 and 91120305), Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20130131120025), the Excellent Youth and Middle Age Scientists Fund of Shandong Province (Grant no. BS2012DX020), the NSFC-JSPS International Cooperation Program (Grant no. 61111140019), the Independent Innovation Foundation of Shandong University, IIFSDU (Grant no. 2011GN062) and the China Postdoctoral Science Foundation (Grant no. 2012M511508).

References

[1] 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, 2009.
[2] A. Cichocki, M. Jankovic, R. Zdunek, S. Amari, Sparse super symmetric tensor factorization, in: Neural Information Processing, Springer, 2008, pp. 781–790.
[3] D. Nion, L. De Lathauwer, A block component model-based blind DS-CDMA receiver, IEEE Trans. Signal Process. 56 (11) (2008) 5567–5579.
[4] D. Nion, L. De Lathauwer, An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA, Signal Process. 88 (3) (2008) 749–755.
[5] Z. He, A. Cichocki, S. Xie, K. Choi, Detecting the number of clusters in n-way probabilistic clustering, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2006–2021.
[6] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1700–1715.

[7] J.D. Carroll, J.J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of Ceckart–Young decomposition, Psychometrika 35 (3) (1970) 283–319.
[8] P.M. Kroonenberg, J. De Leeuw, Principal component analysis of three-mode data by means of alternating least squares algorithms, Psychometrika 45 (1) (1980) 69–97.
[9] R.A. Harshman, M.E. Lundy, Three-way DEDICOM: Analyzing multiple matrices of asymmetric relationships, in: Annual Meeting of the North American Psychometric Society, 1992.
[10] B.W. Bader, R.A. Harshman, T.G. Kolda, Pattern analysis of directed graphs using DEDICOM: an application to enron email, Technical report, Sandia National Laboratories, 2006.
[11] L. De Lathauwer, B. De Moor, J. Vandewalle, et al., A multilinear singular value decomposition, SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1253–1278.
[12] L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, …, r_n) approximation of higher-order tensors, SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1324–1342.
[13] L. De Lathauwer, Decompositions of a higher-order tensor in block terms part i: lemmas for partitioned matrices, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1022–1032.
[14] L. De Lathauwer, D. Nion, Decompositions of a higher-order tensor in block terms part iii: alternating least squares algorithms, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1067–1083.
[15] Y. Panagakis, C. Kotropoulos, G.R. Arce, Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification, IEEE Trans Audio Speech Lang. Process. 18 (3) (2010) 576–588.
[16] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognit. 44 (7) (2011) 1540–1551.
[17] Y.D. Kim, S. Choi, Nonnegative tucker decomposition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, IEEE, 2007, pp. 1–8.
[18] A. Phan, A. Cichocki, Fast and efficient algorithms for nonnegative tucker decomposition, in: Advances in Neural Networks-ISNN 2008, 2008, pp. 772–782.
[19] I. Dhillon, S. Sra, Generalized nonnegative matrix approximations with Bregman divergences, Adv. Neural Inf. Process. Syst. 18 (2006) 283.
[20] A.H. Phan, A. Cichocki, Extended Hals algorithm for nonnegative tucker decomposition and its applications for multiway analysis and classification, Neurocomputing 74 (11) (2011) 1956–1969.
[21] R.A. Harshman, S. Hong, M.E. Lundy, Shifted factor analysis Part I: models and properties, J. Chemom. 17 (7) (2003) 363–378.
[22] M. Mørup, L.K. Hansen, S.M. Arnfred, L.H. Lim, K.H. Madsen, Shift-invariant multilinear decomposition of neuroimaging data, NeuroImage 42 (4) (2008) 1439–1450.
[23] R.B. Buxton, E.C. Wong, L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model, Magn. Reson. Med. 39 (6) (1998) 855–864.
[24] M.I. Sereno, A.M. Dale, J.B. Reppas, K.K. Kwong, J.W. Belliveau, T.J. Brady, B.R. Rosen, R.B. Tootell, Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging, Science 268 (5212) (1995) 889.
[25] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2-Part II. Modeling chromatographic data with retention time shifts, J. Chemom. 13 (3-4) (1999) 295–309.
[26] S. Hong, R.A. Harshman, Shifted factor analysis part II: algorithms, J. Chemom. 17 (7) (2003) 379–388.
[27] S. Hong, R.A. Harshman, Shifted factor analysis Part III: N-way generalization and application, J. Chemom. 17 (7) (2003) 389–399.
[28] M. Mørup, M. Schmidt, Sparse non-negative tensor 2d deconvolution (SNTF2D) for multi channel time–frequency analysis (Technical Report). Technical University of Denmark, 2006.
[29] M. Mørup, M.N. Schmidt, L.K. Hansen, Shift Invariant Sparse Coding of Image and Music Data, 2008.
[30] B. Makkiabadi, F. Ghaderi, S. Sanei, A new tensor factorization approach for convolutive blind source separation in time domain, in: EUSIPCO 2010, 2010.
[31] P. Smaragdis, Convolutive speech bases and their application to supervised speech separation, IEEE Trans. Audio Speech Lang. Process. 15 (1) (2006) 1–12.
[32] A. Cichocki, A.H. Phan, Fast local algorithms for large scale nonnegative matrix and tensor factorizations, IEICE Trans. Fundam. Electron. 92 (2009) 708–721.
[33] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[34] A. Stegeman, Degeneracy in CANDECOMP/PARAFAC and INDSCAL explained for several three-sliced arrays with a two-valued typical rank, Psychometrika 72 (4) (2007) 601–619.
[35] D.D. Lee, S.H. Seung, Algorithms for non-negative matrix factorization, Adv. Neural Inf. Process. Syst. 13 (2000) 556–562.
[36] A. Cichocki, S. Cruces, S. Amari, Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization, Entropy 13 (1) (2011) 134–170.
[37] Q. Wu, L.Q. Zhang, G.C. Shi, Robust feature extraction for speaker recognition based on constrained nonnegative tensor factorization, J. Comput. Sci. Technol. 25 (4) (2010) 783–792.
[38] S. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process. 27 (2) (1979) 113–120.
[39] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies (oasis): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, J. Cogn. Neurosci. 19 (9) (2007) 1498–1507.
[40] E.C. Smith, M.S. Lewicki, Efficient auditory coding, Nature 439 (7079) (2006) 978–982.
[41] A. Hyvarinen, P. Hoyer, E. Oja, Sparse code shrinkage: denoising by nonlinear maximum likelihood estimation, Adv. Neural Inf. Process. Syst. (1999) 473–479.

Qiang Wu received the Ph.D. degree in Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. He is now a lecturer with School of Information Science and Engineering, Shandong University, Shandong, China. His research interests include machine learning, brain-computer interface, medical imaging processing, speech signal processing, neuroscience.

Liqing Zhang received the Ph.D. degree from Zhongshan University, Guangzhou, China, in 1988. He was promoted to full professor position in 1995 at South China University of Technology. He worked as a research scientist in RIKEN Brain Science Institute, Japan from 1997 to 2002. He is now a Professor with Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests cover computational theory for cortical networks, brain signal processing and brain–computer interface, perception and cognition computing model, statistical learning and inference. He has published more than 160 papers in international journals and conferences.

Andrzej Cichocki was born in Poland. He received his M.Sc. (with honors), Ph.D. and Habilitate Doctorate (Dr.Sc.) degrees, all in Electrical Engineering, from the Warsaw University of Technology (Poland). He is the co-author of four international books and monographs (two of them are translated into Chinese): Nonnegative Matrix and Tensor Factorizations (J.Wiley, 2009), Adaptive Blind Signal and Image Processing (J. Wiley, 2002), MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (J. Wiley and Teubner Verlag, 1993/1994) and author or co-author of more than 300 papers. He is Associate Editor of Journal of Neuroscience Methods and IEEE Transactions on Signal Processing. Currently, he is the head of the laboratory for Advanced Brain Signal Processing in Brain Science Institute, RIKEN, Japan.