

Locally Linear Low-rank Tensor Approximation

Alp Ozdemir, Mark A. Iwen and Selin Aviyente

Abstract—Recently, collecting and storing higher order data has become more feasible with the use of methods from multi-linear algebra. High order data usually lies in a low dimensional subspace or manifold along each mode and its intrinsic structure can be revealed by linear methods such as higher order SVD. However, these linear approaches may not capture the local nonlinearities in the data that may occur due to moving sensors or other nonlinearities in the measurements. In this paper, we propose to use a piecewise linear model to better identify the non-linearities in higher order data. The proposed approach decomposes the higher-order data into subtensors and fits a low rank model to each subtensor. The proposed approach is applied to simulated datasets and a video sequence captured across different angles to show its robustness to non-linear structures.

Index Terms—Tensor algebra, higher order SVD, low-rank approximation, manifold learning

I. INTRODUCTION

High dimensional data sets in \mathbb{R}^D are often assumed to be lying near a lower d -dimensional manifold, or subspace, with $d \ll D$. When the data points are located on a linear subspace, this linear structure can be represented by a basis identified via PCA (or SVD). However, if the data points lie near a non-linear manifold, the basis obtained by such linear methods may not encode the data points efficiently [1], [2].

In order to address this issue, many approaches for learning low dimensional manifolds from high dimensional datasets have been developed by the machine learning community (see, e.g., [1]–[4]). For example, Tenenbaum et al. [3], presented an algorithm that consists of embedding the data into a graph and then constructing a d -dimensional manifold by applying multidimensional scaling (MDS) to the graph. Similarly, Roweis and Saul [4] introduced locally linear embedding (LLE) approach by assuming that each data point and its neighbors are lying on a locally linear patch of the manifold. More recently, Allard et al. [1] proposed generating data-dependent multi-scale dictionaries called Geometric Multi-resolution Analysis (GMRA) in order to have a fast and compact encoding system for nonlinear data. Although these approaches are useful for identifying embedded low dimensional manifold for vector type data, they are not directly applicable to high-order data sets, such as tensors which occur in many applications including video, hyperspectral imaging, social and biological networks [5]–[7].

A. Ozdemir (ozdemira@msu.edu) and S. Aviyente (aviyente@egr.msu.edu) are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, 48824, USA.

M.A. Iwen (markiwen@math.msu.edu) is with the Department of Electrical and Computer Engineering and the Department of the Mathematics, Michigan State University, East Lansing, MI, 48824, USA.

This work was in part supported by NSF CCF-1422262, NSF CCF-1218377, NSF DMS-1416752 and NSA H98230-13-1-0275.

Herein we propose a new method for identifying low-dimensional manifold structure from higher-order data. Our proposal is both general and simple: one may simply replace linear methods (e.g., the SVD) with non-linear manifold learning methods when computing tensor factorizations of high-order data. Unsurprisingly, doing so results in superior performance when the data in question has nonlinear, as opposed to linear, low-dimensional structure. Applications of this work include finding low-dimensional approximations to tensor data arising from problems in face recognition, surveillance video analysis, and anomaly detection in social networks.

Previously, these problems involving higher order data have been addressed using higher order SVD (HOSVD). For example, Vasilescu and Terzopoulos [8] extended the eigenface concept to the tensorface by using higher order SVD and taking different modes such as expression, illumination and pose into account. Yang et al. [9] presented 2D-PCA for matrices and used it for feature extraction from face images without converting the images into vectors. Others [10]–[12] used incremental SVD for tensors for background modelling and face recognition. Most of the subspace learning algorithms mentioned above are interested in fitting a low-rank model to data which lies near a *linear* subspace and do not focus on non-linear structures that the datasets possibly have.

More recently, researchers have extended manifold learning approaches to tensors. For example, He et al. [13] extended locality preserving projections [14] to second order tensors and then used it for dimensionality reduction and face recognition in a supervised setting. Dai and Yeung [15] presented generalized tensor embedding methods such as the tensor extension of local discriminant embedding methods [16], neighborhood preserving embedding methods [17], and locality preserving projection methods [14]. Li et al. [18] proposed a discrimination method which preserves local structures in each class, and then extended the algorithm to high-order tensors. However, these methods are mostly limited to learning the optimal linear transformation for supervised classification of high-order data. In contrast, we propose novel unsupervised higher-order manifold learning approaches for summarizing, reducing, and performing anomaly detection within tensor data.

The proposed method will be referred to as locally linear higher order singular value decomposition (LL-HOSVD) which enables us to deal with piecewise linear low-rank structure for high-order data. This algorithm consists of 2 main steps: Decomposing the tensor into lower dimensional subtensors which are expected to have more linear structure, and then applying HOSVD to these subtensors to identify

their underlying low-dimensional linear structure. Moreover, we propose two approaches to decompose the tensor into the subtensors. Finally, we apply the proposed algorithm to two synthetic datasets as well as a video sequence to exhibit its robustness to non-linearities.

II. BACKGROUND

A. Tensor Algebra

A multidimensional array with N modes $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is called a tensor, where x_{i_1, i_2, \dots, i_N} denotes the $(i_1, i_2, \dots, i_N)^{th}$ element of the tensor \mathcal{X} . Vectors obtained by fixing all indices of the tensor except the one that corresponds to n th mode are called mode- n fibers. Basic tensor operations are reviewed below [19], [20].

Tensor norm Norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is square root of sum of the squares of all its elements.

$$\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \dots, i_N}^2}. \quad (1)$$

Tensor inner product The inner product of two same sized tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is sum of the products of their elements.

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \dots, i_N} y_{i_1, i_2, \dots, i_N}. \quad (2)$$

Mode- n product The mode- n product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_n \times \dots \times I_N}$ and a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ is denoted as $\mathcal{Y} = \mathcal{X} \times_n \mathbf{U}$, $(\mathcal{Y})_{i_1, i_2, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} x_{i_1, \dots, i_n, \dots, i_N} u_{j, i_n}$ and is of size $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$.

Tensor matricization Process of reordering the elements of the tensor into a matrix is known as matricization or unfolding. The mode- n matricization of tensor \mathcal{Y} is denoted as $\mathbf{Y}_{(n)}$ and is obtained by arranging mode- n fibers to be the columns of the resulting matrix. Unfolding the tensor $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)}$ along mode- n is equivalent to $\mathbf{Y}_{(n)} = \mathbf{U}^{(n)} \mathbf{X}_{(n)} (\mathbf{U}^{(N)} \otimes \dots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \dots \otimes \mathbf{U}^{(1)})^\top$, where \otimes is the matrix Kronecker product.

Tensor Rank Unlike matrices, which have a unique definition of rank, there are multiple rank definitions for tensors including *tensor rank* and *tensor n -rank*. The *rank* of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_n \times \dots \times I_N}$ is the smallest number of rank-one tensors that form \mathcal{X} as their sum. The *n -rank* of \mathcal{X} is the collection of ranks of mode matrices $\mathbf{X}_{(n)}$ and is denoted as:

$$n\text{-rank}(\mathcal{X}) = (\text{rank}(\mathbf{X}_{(1)}), \text{rank}(\mathbf{X}_{(2)}), \dots, \text{rank}(\mathbf{X}_{(N)})). \quad (3)$$

B. Higher Order Singular Value Decomposition (HOSVD)

Any tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be decomposed as mode products of a core tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and N orthogonal projection matrices $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ which are the left singular vectors of $\mathbf{X}_{(n)}$ [21]:

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_N \mathbf{U}^{(N)} \quad (4)$$

where \mathcal{S} is computed as

$$\mathcal{S} = \mathcal{X} \times_1 (\mathbf{U}^{(1)})^\top \times_2 (\mathbf{U}^{(2)})^\top \dots \times_N (\mathbf{U}^{(N)})^\top. \quad (5)$$

Let $\mathcal{S}_{i_n=\alpha}$ be a subtensor of \mathcal{S} obtained by fixing the n th index to α . This subtensor satisfies following properties:

- all-orthogonality: $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of n , α and β subject to $\alpha \neq \beta$.

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \text{ when } \alpha \neq \beta. \quad (6)$$

- ordering:

$$\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=2}\| \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\| \geq 0 \quad (7)$$

for all possible values of n .

III. LOCAL HIGHER ORDER SINGULAR VALUE DECOMPOSITION

In this section, we present a procedure which provides low n -rank approximation to subtensors of an N th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ to better capture local nonlinearities.

The proposed method starts with decomposing tensor \mathcal{X} into K subtensors $\mathcal{Y}_k \in \mathbb{R}^{I_{1,k} \times I_{2,k} \times \dots \times I_{N,k}}$ with $k \in \{1, 2, \dots, K\}$. \mathcal{Y}_k s are formed by using the following mapping functions f_k s defined on the index sets of \mathcal{X} and \mathcal{Y}_k .

Let f_k be a mapping function from the index set of \mathcal{X} to the index set of \mathcal{Y}_k as $f_k : J_1 \times J_2 \times \dots \times J_N \mapsto J_{1,k} \times J_{2,k} \times \dots \times J_{N,k}$, where $J_n = \{1, 2, \dots, I_n\}$, $J_{n,k} = \{1, 2, \dots, I_{n,k}\}$ with $n \in \{1, 2, \dots, N\}$, and f_k s satisfy $\cup_{k=1}^K J_{n,k} = J_n$ and $J_{n,k} \cap J_{n,l} = \emptyset$ when $k \neq l$ for all $k, l \in \{1, 2, \dots, K\}$. We propose two approaches: direct division and sequential division to obtain the f_k 's and \mathcal{Y}_k 's as explained in the next two sections.

Once \mathcal{Y}_k s are obtained, HOSVD is used to obtain the low n -rank approximation for each \mathcal{Y}_k . This can be achieved by using truncated mode matrices obtained by keeping only a few of the vectors of $\mathbf{U}^{(n,k)}$ corresponding to the highest singular values. For purposes of simplifying the representation, the same n -rank is selected for each \mathcal{Y}_k , $k \in \{1, 2, \dots, K\}$ and this rank is denoted by $R = (r_1, r_2, \dots, r_N)$. Let $\hat{\mathcal{Y}}_k$ be a low n -rank approximation of \mathcal{Y}_k computed as:

$$\hat{\mathcal{Y}}_k = \hat{\mathcal{S}}_k \times_1 \hat{\mathbf{U}}^{(1,k)} \times_2 \hat{\mathbf{U}}^{(2,k)} \dots \times_N \hat{\mathbf{U}}^{(N,k)}, \quad (8)$$

where $\hat{\mathbf{U}}^{(n,k)}$ s are the truncated projection matrices of \mathcal{Y}_k obtained by keeping the first r_n columns of $\mathbf{U}^{(n,k)}$ for $n \in \{1, 2, \dots, N\}$ and $\hat{\mathcal{S}}_k$ is the core tensor

$$\hat{\mathcal{S}}_k = \hat{\mathcal{Y}}_k \times_1 (\hat{\mathbf{U}}^{(1,k)})^\top \times_2 (\hat{\mathbf{U}}^{(2,k)})^\top \dots \times_N (\hat{\mathbf{U}}^{(N,k)})^\top. \quad (9)$$

Therefore,

$$\hat{\mathcal{X}}_{f_k(J_1 \times J_2 \times \dots \times J_N)} = \hat{\mathcal{Y}}_k, \quad (10)$$

and combining all of the subtensors $\hat{\mathcal{Y}}_k$ s by using the inverse mapping functions f_k^{-1} provides piecewise-linear approxima-

tion of \mathcal{X} :

$$\hat{\mathcal{X}} = \sum_{k=1}^K \hat{\mathcal{Y}}_{k, (f_k^{-1}(J_{1,k} \times J_{2,k} \times \dots \times J_{N,k}))}. \quad (11)$$

A pseudo code of the algorithm is given in Algorithm 1.

Algorithm 1 Locally Linear Higher Order SVD

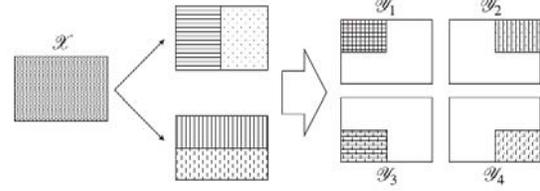
- 1: Input: $\hat{\mathcal{X}}$: tensor, $\mathbf{R} = (r_1, r_2, \dots, r_N)$: the desired local n -rank, $\mathbf{C} = (c_1, c_2, \dots, c_N)$: the desired number of clusters for mode- n fibers.
 - 2: Output: $\hat{\mathcal{X}}$
 - 3: $K \leftarrow \prod_{i=1}^N c_i$
 - 4: $\left[\{\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_K\}, \{f_1, f_2, \dots, f_K\} \right] = \text{direct-division}(\hat{\mathcal{X}}, \mathbf{C})$
or $\text{sequential-division}(\hat{\mathcal{X}}, \mathbf{C})$
 - 5: **for** $k = 1$ to K **do**
 - 6: **for** $n = 1$ to N **do**
 - 7: $(\mathbf{U}^{(n,k)}) \leftarrow \text{SVD}(\mathbf{Y}^{(k,n)})$
 - 8: $(\hat{\mathbf{U}}^{(n,k)}) \leftarrow \text{truncate}(\mathbf{U}^{(n,k)}, r_n)$
 - 9: **end for**
 - 10: $\mathcal{S}_k \leftarrow \mathcal{Y}_k \times_1 (\hat{\mathbf{U}}^{(1,k)})^\top \times_2 (\hat{\mathbf{U}}^{(2,k)})^\top \dots \times_N (\hat{\mathbf{U}}^{(N,k)})^\top$
 - 11: $\hat{\mathcal{Y}}_k \leftarrow \mathcal{S}_k \times_1 \hat{\mathbf{U}}^{(1,k)} \times_2 \hat{\mathbf{U}}^{(2,k)} \dots \times_N \hat{\mathbf{U}}^{(N,k)}$,
 - 12: **end for**
 - 13: $\hat{\mathcal{X}} = \sum_{k=1}^K \hat{\mathcal{Y}}_{k, (f_k^{-1}(J_{1,k} \times J_{2,k} \times \dots \times J_{N,k}))}$
-

A. Direct Division

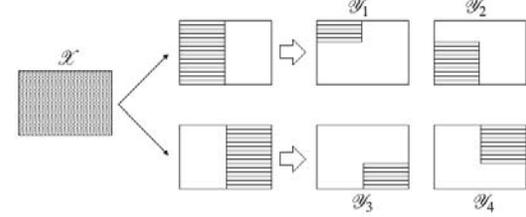
In this approach, tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is first unfolded across each mode yielding $\mathbf{X}_n \in \mathbb{R}^{I_n \times \prod_{j \neq n} I_j}$ whose columns are the mode- n fibers of \mathcal{X} . For each mode, the mode- n fibers are partitioned into c_n non-overlapping clusters. We made use of diffusion maps to partition the mode- n fibers as follows [22]: Let $\mathbf{x}_{n,i}$ be the i th mode- n fiber of \mathcal{X} . First, each $\mathbf{x}_{n,i}$ is connected to its k -nearest neighbors with weights $W_{i,j} = K(\mathbf{x}_{n,i}, \mathbf{x}_{n,j}) = e^{-\|\mathbf{x}_{n,i} - \mathbf{x}_{n,j}\|^2 / \epsilon_i \epsilon_j}$, where ϵ_i is the Euclidean distance between $\mathbf{x}_{n,i}$ and its $k/2$ -nearest neighbors, to obtain the weighted graph on the mode- n fibers. We then use the METIS algorithm [23] to partition the constructed graph for clustering the fibers. This process is applied for each mode separately. Then, Cartesian product of the fiber labels coming from different modes yields $K = \prod_{i=1}^N c_n$ subtensors \mathcal{Y}_k . This approach is illustrated in Fig. 1a for 2-way tensors.

B. Sequential Division

In this approach, non-overlapping tensors \mathcal{Y}_k s are obtained by applying clustering to the fibers across each mode and generating non-overlapping tensors N times repeatedly as follows. First mode-1 fibers of \mathcal{X} are grouped into c_1 clusters using the same two step procedure as in approach-1 through a weighted graph constructed from the fibers and then partitioning using METIS. Then, mode-2 fibers of each of the newly created tensors are clustered into c_2 clusters separately which yields $c_1 \times c_2$ subtensors. This procedure is applied N times by clustering the fibers of different modes at each step and $K = \prod_{i=1}^N c_n$ subtensors \mathcal{Y}_k are obtained (Fig. 1b). However, in this procedure, it is not necessary to start from the mode-1 fibers and choosing different ordering of the modes yields $N!$ different possible decompositions, \mathcal{Y}_k .



(a) Direct division



(b) Sequential division

Fig. 1. Illustration of creation of 4 non-overlapping tensors \mathcal{Y}_k from a 2-mode tensor \mathcal{X} by (a) direct division and (b) sequential division.

IV. RESULTS

In this section, we will evaluate the effectiveness of the locally linear higher order SVD with direct division (LL-HOSVD(DD)) and sequential division (LL-HOSVD(SD)) approaches for revealing the non-linear structure of tensor type data.

A. Translating Subspaces

In this experiment, we generate two point clouds each containing 100 Gaussian random variables with zero mean and identity covariance matrix in \mathbb{R}^{20} and the subspaces in which the point clouds live are orthogonal to each other in \mathbb{R}^{100} . The first point cloud is static whereas the second one is translating in a subspace that is orthogonal to the first subspace. After the time point $t = 20$ the dynamic subspace starts to overlap with the static subspace and at $t = 30$ the dynamic cloud starts to move back to its original place. A 3-mode tensor $\mathcal{X} \in \mathbb{R}^{100 \times 200 \times 60}$ is created to represent the data samples, where the third mode is time. Low n -rank approximations of \mathcal{X} are computed by HOSVD and the proposed approaches LL-HOSVD(DD) and LL-HOSVD(SD). Various values of n are used in the experiments and the cluster number along each mode is chosen as $C = (4, 4, 4)$ (Table I). Since the LL-HOSVD(SD) has $N! = 6$ possible cluster combinations, we just reported the cluster combination with the lowest mean squared error (MSE). As seen in Table-I, both of the proposed approaches give better approximation than HOSVD in terms of MSE, and the data clouds in the resulting low rank tensor obtained by the proposed approaches are better separated than those obtained by HOSVD (Figure 2). Since LL-HOSVD(SD) iteratively divides the subtensors when creating \mathcal{Y}_k s, it tends to provide finer approximation. As the rank increases, the approximation gets better as expected.

TABLE I
AVERAGE MSE FOR THE RECONSTRUCTED 3-WAY TENSOR $\mathcal{X} \in \mathbb{R}^{100 \times 200 \times 60}$ FOR MOVING SUBSPACES BY HOSVD, LL-HOSVD(DD) AND LL-HOSVD(SD) APPROACHES AT VARYING n -RANK OVER 25 TRIALS.

	$R = (3, 3, 3)$	$R = (5, 5, 5)$	$R = (7, 7, 7)$	$R = (9, 9, 9)$
HOSVD	0.1131	0.1006	0.0943	0.0885
LL-HOSVD(DD)	0.1029	0.0926	0.0792	0.0662
LL-HOSVD(SD)	0.0838	0.0584	0.0407	0.0285

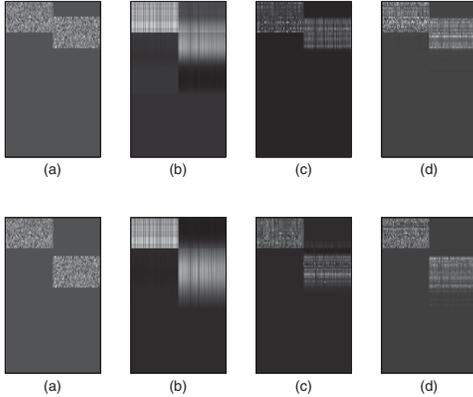


Fig. 2. Sample outputs for two different time points: (a) original image, (b) HOSVD, (c) LL-HOSVD(DD), (d) LL-HOSVD(SD)

B. Rotating Subspaces

In this section, we generate two point clouds each containing 100 Gaussian random variables with zero mean and a covariance matrix corresponding to 1st order Gauss-Markov process in \mathbb{R}^{20} and the subspaces in which the point clouds live are orthogonal to each other in \mathbb{R}^{100} . The first point cloud is static whereas the second one is rotating at a constant speed. To be able to adjust the speed of the rotation, the rotation matrix A is constructed as $A = I_{10 \times 10} \otimes \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$. At $t = 10$, the angle of the rotation increased to $\theta = \pi/60$ from $\theta = \pi/120$. A 3-mode tensor $\mathcal{X} \in \mathbb{R}^{100 \times 200 \times 80}$ is created to represent data samples, where the third mode is time. Low n -rank approximations of \mathcal{X} are computed by HOSVD and the proposed approaches LL-HOSVD(DD) and LL-HOSVD(SD). The various n -ranks for $\hat{\mathcal{X}}$ are used in the experiments and the cluster number along each mode is chosen as $C = (4, 4, 4)$ (Table II). Similar to previous simulations, we reported one of the LL-HOSVD(SD) results with the lowest MSE. As seen in Table-II, both of the proposed approaches gives better approximation than HOSVD in terms of MSE.

C. PIE Dataset

In this experiment, a 3-mode tensor $\mathcal{X} \in \mathbb{R}^{122 \times 160 \times 138}$ is created from PIE dataset [24]. The tensor contains 138 images from 6 different yaw angles and varying illumination

TABLE II
AVERAGE MSE FOR THE RECONSTRUCTED 3-WAY TENSOR $\mathcal{X} \in \mathbb{R}^{100 \times 200 \times 60}$ FOR ROTATING SUBSPACES BY HOSVD, LL-HOSVD(DD) AND LL-HOSVD(SD) APPROACHES AT VARYING n -RANK OVER 25 TRIALS.

	$R = (3, 3, 3)$	$R = (5, 5, 5)$	$R = (7, 7, 7)$	$R = (9, 9, 9)$
HOSVD	0.1016	0.0896	0.0786	0.0685
LL-HOSVD(DD)	0.685	0.0540	0.0474	0.0432
LL-HOSVD(SD)	0.0493	0.0231	0.0137	0.0084

conditions collected from a subject where each image is converted to gray scale and downsampled to 122×160 . n -rank for \mathcal{X} is determined empirically as $(20, 25, 15)$ for this experiment and the cluster number along each mode is chosen as $C = (4, 4, 4)$. Similar to previous simulations, we reported the LL-HOSVD(SD) results with the lowest MSE. As seen in Fig. 3, the proposed approaches provide more details for the faces with reduced MSE while conventional HOSVD gives blurry approximations.

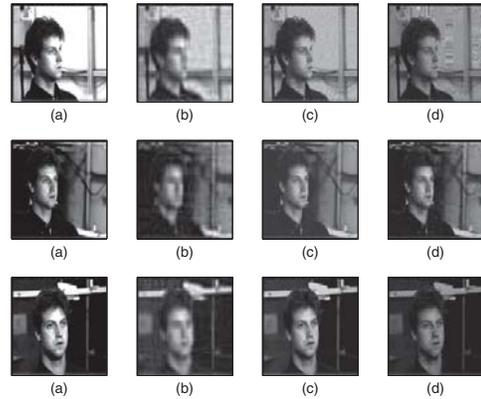


Fig. 3. Frames corresponding to 3 different yaw angles obtained from approximated low n -rank tensor: (a) original image, (b) HOSVD, MSE = 439.0140, (c) LL-HOSVD(DD), MSE = 140.6469, (d) LL-HOSVD(SD), MSE = 378.3899

V. CONCLUSIONS

In this study, we introduced a new low-rank tensor approximation technique which learns the underlying nonlinear structure by fitting locally linear low rank subtensors. We also proposed two approaches to decompose a tensor into its subtensors. The proposed framework is evaluated by applying it to a set of simulated datasets and a video from different angles. Future work will consider automatic selection of parameters in the algorithm such as the number of clusters along each mode and the appropriate rank along each mode. Combining the algorithm with the multiscale structure of GMRA will enable us to obtain a multi-resolution tree structure for high order datasets. Adapting the method to dynamic tensors for identifying structural changes to the tensor in time will also be considered.

REFERENCES

- [1] W. K. Allard, G. Chen, and M. Maggioni, "Multi-scale geometric methods for data sets ii: Geometric multi-resolution analysis," *Applied and Computational Harmonic Analysis*, vol. 32, no. 3, pp. 435–462, 2012.
- [2] Z. Zhang, J. Wang, and H. Zha, "Adaptive manifold learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 2, pp. 253–265, 2012.
- [3] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [5] D. Letexier, S. Bourennane, and J. Blanc-Talon, "Nonorthogonal tensor matricization for hyperspectral image filtering," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 1, pp. 3–7, 2008.
- [6] T.-K. Kim and R. Cipolla, "Canonical correlation analysis of video volume tensors for action categorization and detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 8, pp. 1415–1428, 2009.
- [7] F. Miwakeichi, E. Martinez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing eeg data into space–time–frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.
- [8] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Computer Vision/ECCV 2002*. Springer, 2002, pp. 447–460.
- [9] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 1, pp. 131–137, 2004.
- [10] W. Hu, X. Li, X. Zhang, X. Shi, S. Maybank, and Z. Zhang, "Incremental tensor subspace learning and its applications to foreground segmentation and tracking," *International Journal of Computer Vision*, vol. 91, no. 3, pp. 303–327, 2011.
- [11] A. Sobral, C. G. Baker, T. Bouwmans, and E.-h. Zahzah, "Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction," in *Image Analysis and Recognition*. Springer, 2014, pp. 94–103.
- [12] J. Li, G. Han, J. Wen, and X. Gao, "Robust tensor subspace learning for anomaly detection," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 89–98, 2011.
- [13] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Advances in neural information processing systems*, 2005, pp. 499–506.
- [14] X. Niyogi, "Locality preserving projections," in *Neural information processing systems*, vol. 16. MIT, 2004, p. 153.
- [15] G. Dai and D.-Y. Yeung, "Tensor embedding methods," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 21, no. 1. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006, p. 330.
- [16] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 846–853.
- [17] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1208–1213.
- [18] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 38, no. 2, pp. 342–352, 2008.
- [19] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [20] V. De Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1084–1127, 2008.
- [21] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [22] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, pp. 7426–7431, 2005.
- [23] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.
- [24] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615–1618, 2003.