A GENERAL FRAMEWORK FOR ROBUST HOSVD-BASED INDEXING AND RETRIEVAL WITH HIGH-ORDER TENSOR DATA

Qun Li, Student Member, IEEE, Xiangqiong Shi and Dan Schonfeld, Fellow, IEEE

University of Illinois at Chicago, IL 60607, Email: {qli27, xshi4, dans}@uic.edu

ABSTRACT

In this paper, we first present a theorem that HOSVD-based representation of high-order tensor data provides a robust framework that can be used for a unified representation of the HOSVD of all subtensors. We then propose a general algorithm for robust indexing and retrieval of multiple motion trajectories obtained from a multi-camera system. Guided by our theorem, the unitary transformation matrices of a subset of the original unitary matrices can be very well approximated by a subset of unitary matrices. Simulation results are finally used to illustrate the robustness and efficiency of the proposed approach to multiple trajectory indexing and retrieval from multi-camera systems.

Index Terms—Multi-Camera System, Tensor Tucker Decomposition, HOSVD, Trajectory Indexing, Trajectory Retrieval.

1. INTRODUCTION

In the field of motion trajectory indexing and retrieval where the data set is formed into high-order tensor, high-order singular value decomposition (HOSVD), serving as a main method for compact representation and dimensionality reduction has received much attention. Pioneered in [1][2], the x- and y-location information of multiple motion trajectories at each frame is concatenated as one feature vector which enables the representation of image-as-matrix and hence video-as-tensor. In [3], Xiang et al. represent motion information in a video database as a 3rd-order tensor to which they apply HOSVD to extract both intrinsic and common characteristics of data. The proposed Geometrical Multiple-Trajectory Indexing and Retrieval (GMIR) Algorithm has demonstrated its promising performance in analysis of multiple-object motion trajectory data. We thus demonstrate analytically and experimentally that the proposed HOSVD-based representation can handle flexible query structure consisting of an arbitrary number of objects, cameras and modalities.

2. THEORETICAL FOUNDATION

To extend the work presented in [4] to higher order tensors, we propose our theorem.

Lemma 1. [4, Thm. 3.3] Given a sequence of unitary matrices \( U_k \in R^{L_1 \times L_2 \times \ldots \times L_M} \), and a tensor \( X \in R^{L_1 \times L_2 \times \ldots \times L_M} \), the function \( f(X) = \|X - \hat{X}\|_F^2 \), where \( rank_k(\hat{X}) = L_k \), is minimized, when \( \hat{X} \in R^{L_1 \times L_2 \times \ldots \times L_M} \) by \( \hat{X} = \mathcal{X} \prod_{k=1}^{M} x_k (U_k U_k^T) \).

Lemma 2. [4, Thm. 3.4] For a given tensor \( X \in R^{L_1 \times L_2 \times \ldots \times L_M} \), minimizing \( f(U_{k=1}^M) = \|X - \mathcal{X} \prod_{k=1}^{M} x_k (U_k U_k^T)\|_F^2 \) is equivalent to maximizing \( g(U_{k=1}^M) = \|X \prod_{k=1}^{M} x_k U_k^T\|_F^2 \).

Theorem 1. \( X \in R^{L_1 \times L_2 \times \ldots \times L_M} \), \( \{U_1^*, U_2^*, \ldots, U_M^*\} = \arg \min_{\{U_1, U_2, \ldots, U_M\}} \sum_i \|X_i - \mathcal{X} \prod_{k=1}^{M} x_k (U_k U_k^T)\|_F^2 \), \( \{i_1, i_2, \ldots, i_n\} \subset \{1, 2, \ldots, M\} \), is equivalent to the \( \{U_1^*, U_2^*, \ldots, U_M^*\} \) tuple of \( \{U_1^*, U_2^*, \ldots, U_M^*\} = \arg \min_{\{U_1, U_2, \ldots, U_M\}} \|X - \mathcal{X} \prod_{k=1}^{M} x_k U_k^T\|_F^2 \), where \( \mathcal{X} \in R^{L_1 \times L_2 \times \ldots \times L_M} \), \( \{i_{n+1}, \ldots, i_M\} = \{1, 2, \ldots, M\} \), \( U_{i_k} \in R^{L_1 \times L_2 \times \ldots \times L_M} \), where \( i_k \in \{i_1, \ldots, i_n\}, L_{i_k} \geq L_{i_k}' \) have orthogonal columns, while \( U_{i_k}^* \in R^{L_1 \times L_2 \times \ldots \times L_M} \), \( i_k \in \{i_{n+1}, \ldots, i_M\} \), are orthogonal matrices.

Proof. Define \( C_1 = \mathcal{X} \prod_{k=1}^{M} x_k U_k^T \) and hence \( \mathcal{X} \prod_{k=1}^{M} x_k U_k^T = C_1 \prod_{k=M+1}^{M} x_k U_k^T \).

For \( l = 1 \), which means we only separate one term from \( \mathcal{X} \prod_{k=1}^{M} x_k U_k^T \), \( C_{1(M)} \) is the unfolding of \( C_1 \) over the \( i_k \)th
mode, according to Lemma 2,

\[ (U_1^*, ..., U_M^*) = \arg \min \left( \prod_{k=1}^M \| x_k U_k \|_F^2 \right) \]

\[ = \arg \max \left( \prod_{k=1}^M \| x_k U_k \|_F^2 \right) \]

\[ = \arg \max \left( \| C_1 \|_F \right) \]

\[ = \arg \max \left( \| U_{iM}^T C_{iM} \|_F \right) \]

\[ = \arg \max \left( \| U_{iM}^T \|_F \right) \]

\[ = \arg \max \left( \| X \|_F \right) \]

Suppose this establishes for \( l = m \), that is, \( (U_1^*, ..., U_M^*) = \arg \max \left( \| X \|_F \right) \).

When \( l = m + 1 \), we have

\[ (U_1^*, ..., U_M^*) = \arg \min \left( \prod_{k=1}^M \| x_k U_k \|_F^2 \right) \]

\[ = \arg \max \left( \prod_{k=1}^M \| x_k U_k \|_F^2 \right) \]

\[ = \arg \max \left( \| C_m \|_F \right) \]

\[ = \arg \max \left( \| U_{iM}^T \|_F \right) \]

\[ = \arg \max \left( \| X \|_F \right) \]

Similarly, we arrive at

\[ (U_1^*, ..., U_M^*) = \arg \min \left( \prod_{k=1}^M \| x_k U_k \|_F^2 \right) \]

This completes the proof.

\[ \Box \]

As we know, when we apply HOSVD to a \( N^{th} \)-order tensor, we obtain \( N \) unitary matrices. In the field of indexing and retrieval or classification applications dealing with tensor data, most SVD-based algorithms obtain these matrices by applying SVD to the whole data tensor. Obviously, the flexibility of their algorithms regarding flexible dimensional tensor data will suffer significantly. Whenever the dimension of the data tensor changes, by which we mean obtaining lower dimensional tensor from the original whole data tensor, recalculation of SVD is required to obtain a new set of unitary matrices. Guided by our theorem, unitary matrices of the \( M^{th} \)-order subtensor, where \( M \leq N \), can be very well approximated by the corresponding subset of all the \( N \) unitary matrices, hence avoid recalculation. For instance, in the following experiment, the video database is represented as a
4-D tensor $\mathcal{T}$. By HOSVD we obtain four projection matrices $U_1, U_2, U_3$ and $U_4$ representing the main axes of variation along each axis in a 4-D system. To be specific, $U_1$ spans the space of spatial-temporal multiple trajectory, $U_2$ spans the space of objects, $U_3$ spans the space of sets of multiple trajectories, while $U_4$ spans the space of cameras. At the same time, $U_1, U_2$, and $U_3$ can be also used as projection matrices of the 3-D multiple trajectory tensor from one camera. In fact, guaranteed by our theorem, this applies to all unitary transformations such as Fourier Transform and HOSVD.

3. MULTI-CAMERA MULTI-TRAJECTORY INDEXING AND RETRIEVAL ALGORITHM

Suppose the video database can be formed into a $N^{th}$-order tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, where $I_1, I_2, \ldots, I_N$ may represent different modality of the camera, called different factors for future reference. And the query tensor $\mathcal{T}_{\text{Query}}$ may be with arbitrary selected factors $I_{i_1}, I_{i_2}, \ldots, I_{i_M}$, where $M \leq N$.

3.1. The Indexing Procedure

Usually, in real retrieval application, factors of the query are known. Even if they are unknown, we can easily know them by detecting dimensionality of the query tensor in that different factor has different dimensionality generally.

Step 1: Applying HOSVD to $N^{th}$-order tensor $\mathcal{T}$, we generate data-dependant unitary projection matrices $U_1, U_2, \ldots, U_N$ and the core tensor $\mathcal{S}$:

$$\mathcal{T} = \mathcal{S} \prod_{k=1}^{N} \times_{i_k} U_k.$$  

Step 2: Projecting $N^{th}$-order tensor $\mathcal{T}$ onto certain projection matrices corresponding to the factors of the query, we can perform the retrieval process in the lower-dimensional subspace.

$$\hat{\mathcal{T}} = \mathcal{T} \prod_{k=1}^{M} \times_{i_k} U_k^T.$$  

Step 3: Slicing $\hat{\mathcal{T}}$ into index tensors $\mathcal{T}_{\text{Index}} \in \mathbb{R}^{I_{i_1} \times I_{i_2} \times \ldots \times I_{i_M}}$ where the number of the index tensors is $I_{i_{M+1}} \times I_{i_{M+2}} \times \ldots \times I_{i_N}$.

To take advantage of the unitary transformation, when order of the query tensor $M$ is greater than half of that of the database tensor $N$, instead of projecting the original $N^{th}$-order tensor, we project the core tensor $\mathcal{S}$ onto unitary matrices corresponding to the factors not included in the query, that is, $I_{i_{M+1}}, I_{i_{M+2}}, \ldots, I_{i_N}$, and hence considerably save computation time.

3.2. The Retrieval Procedure

Corresponding to the factors of the query tensor, we retrieve those whose $D_F$ is less than a certain threshold.

$$D_F = \| \mathcal{T}_{\text{Index}} - \mathcal{T}_{\text{QueryIndex}} \|^2_F.$$  

4. EXPERIMENTAL RESULTS

We have implemented the proposed algorithm on an Intel P Dual-Core 1.86GHz Laptop with 1G RAM, using Matlab without code optimization. Based on the 2nd CAVIAR data set, we get a tensor of size $200 \times 2 \times 47 \times 2$.

4.1. Complete query—two trajectories from two cameras

The query is a $200 \times 2 \times 2$-tensor, that is, combination of multiple trajectory matrices from two camera. The precision and recall curve in this case is depicted in Fig. 1(a). Retrieval results are shown in Fig. 2.

4.2. Partial query—single trajectory from two cameras

In the second case, the query of size $200 \times 2$, is formed of single trajectories from two cameras. The retrieval results and precision and recall curve in this case are shown in Fig. 3, Fig. 1(b) respectively.
5. CONCLUSION

High-order singular value decomposition (HOSVD) has served as a main approach for compact representation and dimensionality reduction in the field of motion trajectory indexing and retrieval where the data set is formed into high-order tensors. In this paper, we first presented a theorem to show that given an $N^{th}$-order tensor $T$ with $N$ unitary matrices obtained by HOSVD, for any arbitrary $M^{th}$-order subtensor of $T$, where $M \leq N$, the set of unitary matrices of the subtensor via applying HOSVD to this subtensor can be very well approximated by the corresponding subset of all the $N$ unitary matrices. Based on this theorem, we then propose a general framework for robust indexing and retrieval with high-order tensor data. In application, the tensor data can be of arbitrary high order corresponding to color, multi-spectrum and multi-modality. We showed that the HOSVD-based representation provides a unified framework for the HOSVD representation of all subtensors. We also demonstrated experimentally that the proposed approach can be used to handle query structure consisting of an arbitrary number of objects and cameras. Furthermore, simulation results were used to illustrate the superior performance of the proposed approach to multiple trajectory indexing and retrieval from multi-camera systems.

Future work will focus on solving the main problem remaining related to applications where the correspondence between the tensor structure in the query and database is unknown. For instance, we may not always know the correspondence between the motion trajectories associated with a specific object obtained from a particular camera in the query and database.

Acknowledgement

The authors would like to give their thanks to Prof. Dacheng Tao in the School of Computer Engineering, Nanyang Technological University, for discussions that led to the theoretical foundation in this paper.

6. REFERENCES


