VIEW-INVARIANT TENSOR NULL-SPACE REPRESENTATION FOR MULTIPLE MOTION TRAJECTORY RETRIEVAL AND CLASSIFICATION

Xu Chen, Dan Schonfeld and Ashfaq Khokhar

Department of Electrical and Computer Engineering, University of Illinois at Chicago
{xchen27, dans, ashfaq}@uic.edu

ABSTRACT

In this paper, we propose a novel general framework for tensor based null space affine invariants, namely, *tensor null space invariants* (TNSI) with a linear classifier for high order data classification and retrieval. We first derive TNSI, which is perfectly invariant to multidimensional affine transformations due to camera motions for multiple motion trajectories in consecutive motion events. We subsequently propose an efficient classification and retrieval system relying on TNSI for archiving and searching motion events consisting of multiple motion trajectories. The simulation results demonstrate superior performance of the proposed systems.

***Index Terms*—** Classification, multilinear, null space, retrieval, tensor, trajectory.

1. INTRODUCTION

Among affine view-invariance systems, majority of them represent affine view-invariance in a single dimension [1], thus limiting the system to only single object motion based queries and single dimension affine view-invariance. In many applications, it is not only the individual movement of an object that is of interest, but also the motion patterns that emerge while considering synchronized or overlapped movements of multiple objects. For example, in sports video analysis, we are often interested in a group activity involving motion activity of multiple players, rather than the activity of each individual player. Moreover, due to camera movement, same motion trajectory has completely different representations from different viewing angles. Therefore, a highly efficient classification and retrieval system which is invariant to multidimensional affine transformations is highly desirable.

In this paper, we propose a novel fundamental mathematical framework for tensor null space invariants and use this framework for the important application of view-invariant classification and retrieval of motion events involving multiple motion trajectories.[1] *Our main contributions are as follows: (i) we introduce tensor null space invariants (TNSI) which are perfect affine invariants in a multi-dimensional space, (ii) we demonstrate robustness and superiority of TNSI as a powerful tool in classification and retrieval of high-dimensional data over traditional approaches.*

[1]Tensor representation in a finite-dimensional space is often referred to as multilinear algebra.

The rest of the paper is organized as follows: In Section 2, the mathematical formulation and two important properties of tensor null space invariants (TNSI) are introduced. In Section 3, an efficient classification and retrieval system based on TNSI is presented. Section 4 presents simulation results using real-life trajectories from the Australian Sign Language database. Finally, in Section 5, we provide a brief summary of the results.

2. TENSOR NULL SPACE INVARIANTS

Let us denote the tensor $A \in R^{I_1 \times I_2 \dots I_{N-1} \times I_N}$ as the multi-dimensional data. Elements of A are denoted as $a_{i_1 i_2 \dots i_N}$. As in [2], a generalization of the product of two matrices is the product of a tensor and a matrix. The *mode-n product* of a tensor $A \in R^{I_1 \times I_2 \dots \times I_n \dots I_{N-1} \times I_N}$ by a matrix $U \in R^{I_n \times J_n}$, denoted by $A \times_n U$, is a tensor $B \in R^{I_1 \times \dots I_{n-1} \times J_n \times I_{n+1} \dots \times I_N}$ whose entries are:

$$(A \times_n U)_{i_1 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n} a_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} u_{i_n j_n} \quad (1)$$

The mode-n product $B = A \times_n U$ can be computed via the matrix multiplication $B_{(n)} = U A_{(n)}$, followed by a re-tensorization to undo the mode-n flattening.

As in [1], let $Q_i = (x_i, y_i)$ be a single 2-D point, $i = 0, 1, \dots, n-1$, among n ordered non-linear points in R^2. Consider the following arrangement of the n 2-D points in a $3 \times n$ matrix M:

$$M = \begin{pmatrix} x_0 & x_1 & \dots & x_{n-1} \\ y_0 & y_1 & \dots & y_{n-1} \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad (2)$$

the null space H^{n-3} can be represented as:

$$H^{n-3} = \{q = (q_0, q_1, \dots, q_{n-1})^T, i.e. Mq = (0,0,0)^T\} \quad (3)$$

Applying the affine transformation T on the matrix $M_1 = TM$, the null spaces of M_1 and M are identical as described in [1]. Similarly, applying the affine transformation T_m, T_n on the mth, nth unfolding of the multi-dimensional data M, respectively, if the resulting tensor null space Q is invariant in both dimensions, the it is referred to as *mode-m,n invariant*. WLOG, let us derive the mathematical formulation of the mode-1,2,3 invariant tensor Q for three dimensional data $M \in R^{I_1 \times I_2 \times I_3}$. To be rotation invariant, we have:

$$M_{(1)} \times Q_{(3)} = 0\ , M_{(2)} \times Q_{(2)} = 0\ , M_{(3)} \times Q_{(1)} = 0\ , \quad (4)$$

where $M_{(1)}, M_{(2)}, M_{(3)}$ are the unfolding of the three order tensor M into matrices with the dimension $I_2 I_3 \times I_1$, $I_1 \times I_3 I_2$ and $I_1 \times I_2 I_3$, respectively, and $Q_{(3)}, Q_{(2)}, Q_{(1)}$ are the corresponding unfoldings of the tensor Q. On the right side of (4), 0 represents the corresponding null tensor.

Let us assume the translation vector as $T_i = [t_i, \ldots, t_i]$, the translation matrix as $T = [T_1, \ldots, T_N]'$. For example, for motion trajectories, $T = [T_1, T_2]'$, where T_1 represents the shift of all the coordinates in x dimension and T_2 represents the shift of all the coordinates in y dimension. Now we shall derive the condition on the TNSI Q to guarantee the invariance of translation for tensor M. To derive the translation in the dimension I_n, we should unfold the tensor M into matrix $M_{(n)} = [m_1, ..., m_N]'$ with the dimension $I_1 I_2 \times \ldots I_{n-1} I_{n+1} \ldots I_N \times I_n$. According to the definition of tensor null space invariants:

$$\begin{aligned}(RM_{(n)} + T)Q = [Rm_1 + T_1, ..., Rm_N + T_N]'Q = \\ = [Rm_1, ..., Rm_N]'Q + [T_1, ..., T_N]'Q = 0 \qquad (5)\end{aligned}$$

Due to invariance of rotation, $[Rm_1, ..., Rm_N]'Q = 0$. Thus,

$$\begin{aligned}[T_1, ..., T_N]'Q = [T_1 Q, ..., T_N Q]' = [\sum t_1 q_1, ..., \\ \sum t_N q_N] = [t_1, ..., t_N] \sum q_i = 0 \qquad (6)\end{aligned}$$

Since $[t_1, ..., t_N]$ can be arbitrary, the condition for the equation (6) to hold true is that each column of the unfolding of TNSI should sum up to zero, namely, $\sum q_i = 0$. Therefore, we obtain the condition for invariance of translation for the mode-1,2,3 invariant tensor Q:

$$\sum Q_{(1)} = 0, \sum Q_{(2)} = 0, \sum Q_{(3)} = 0 \qquad (7)$$

Combining equations (4) and (7), we can solve the TSNI Q subject to the mode-1,2,3 affine view-invariant. If the order of the tensor M is 2, it is easy to show that the condition boils down to the Stiller's one dimensional null space invariants [1]. It is also easy to extend the result to the case of the Nth order tensor with mode-$I_a, \ldots, I_k$ affine view-invariant.

Lemma 1: For the TNSI tensor Q satisfying n dimensional affine view-invariant, Q should be a n+1 order tensor, where the first n dimension are determined by the size of the data M, the $n+1$th dimension are determined by the number of the basic solutions. Specifically, if the dimension of the N order tensor M $I_1 \times I_2 \times \ldots \times I_n \times I_{n+1} \times \ldots \times I_N$ is subject to n dimensional affine transformation in the first n dimension $I_1 \times I_2 \times \ldots \times I_n$, the corresponding n+1 order tensor TNSI $Q_{W_1 \times W_2 \times \ldots \times W_{n+1}}$ have the dimension as follows:

$$W_1 = I_2 I_3 \ldots I_N\ , W_2 = I_1 I_3 \ldots I_N\ , \ldots\ , \quad (8)$$

$$W_n = I_1 \ldots I_{n-1} I_{n+1} \ldots I_N\ , W_{n+1} = t - rank(A)\ , \quad (9)$$

where A is the coefficient matrix in (10).

Remark: W_{n+1} is determined by the number of basic solutions, namely, the rank of the coefficient matrix A in the equation (10) according to *rank theorem*. In the n+1 order tensor Q, the first n order is determined by the corresponding affine view-invariant. The $n+1$th dimension is the space for the number of basis, namely, the number of basic solutions for the underdetermined linear equations. Theoretically, it fits perfectly with the theory for one dimensional affine view invariant [1], TNSI reduce to a matrix, where the column space is spanned by null space operators. Since the equations (4) and (7) are all linear equations, we vectorize the unknown elements of Q and rewrite all the equations in the matrix form as:

$$\begin{pmatrix} m_{11} & m_{12} & \ldots & m_{1,t-1} & m_{1t} \\ m_{21} & m_{22} & \ldots & m_{2,t-1} & m_{2t} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{s1} & m_{s2} & \ldots & m_{s,t-1} & m_{st} \end{pmatrix} \begin{pmatrix} q_1^{(i)} \\ q_2^{(i)} \\ \ldots \\ q_t^{(i)} \end{pmatrix} = 0\ , \quad (10)$$

where in (10), m_{ij} represents the coefficient formed by the high order data according to (4) and (7) and $q_i = [q_1^{(i)}, \ldots q_t^{(i)}]$ represents the ith basis for tensor null space invariants. The TNSI is spanned by all the basis $q_1, ..., q_k$ and obtained by computing all the basic solutions from (10) and doing retensorization.

Lemma 2: For $s < i \le t$, relying on the equation (10), the constructive solutions for the ith basis for TNSI is given by:

$$q_1^{(i)} = det \begin{pmatrix} m_{12} & m_{13} & \ldots & m_{1s} & m_{1,i} \\ m_{22} & m_{23} & \ldots & m_{2s} & m_{2,i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{s2} & m_{s3} & \ldots & m_{ss} & m_{s,i} \end{pmatrix}\ , \quad (11)$$

$$q_2^{(i)} = -det \begin{pmatrix} m_{11} & m_{13} & \ldots & m_{1s} & m_{1,i} \\ m_{21} & m_{23} & \ldots & m_{2s} & m_{2,i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{s1} & m_{s3} & \ldots & m_{ss} & m_{s,i} \end{pmatrix}\ , \quad (12)$$

$$\ldots, \quad (13)$$

$$q_s^{(i)} = (-1)^s det \begin{pmatrix} m_{11} & m_{12} & \ldots & m_{1,s-1} & m_{1,i} \\ m_{21} & m_{22} & \ldots & m_{2,s-1} & m_{2,i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{s1} & m_{s2} & \ldots & m_{s,s-1} & m_{3,i} \end{pmatrix}\ , \quad (14)$$

$$q_k^{(i)} = 0, \forall k = s+1, ..., t\ . \quad (15)$$

Remark: Lemma 2 can be easily proved by the basic knowledge from linear algebra. According to lemma 2, the TNSI is *uniquely* determined by the original data M and the dimensions that = are subject to affine invariance.

3. CLASSIFICATION AND RETRIEVAL ALGORITHMS

We align each trajectory as two rows in a matrix according to x and y coordinates, and the number of rows of a matrix is set to be twice the number of the objects in the motion event under analysis.

$$M = (M_{i,j})_{i=1,2,\ldots,2J;j=1,2,\ldots,P} \ , \quad (16)$$

In the above equation, P denotes the temporal length of normalized trajectories, J represents the number of trajectories within one motion event. Finally, multiple trajectory matrices are aligned in the direction orthogonal to the plane spanned by them, and form a three dimensional matrix, or tensor. We refer to it as *Motion Event Tensor* T. as shown in Fig. 1.

$$T = (T_{i,j,k})_{i=1,2,\ldots,2J;j=1,2,\ldots,P;k=1,2,\ldots,K} \ , \quad (17)$$

where K is the number of motion event samples (trajectory video clips). Once we have generated the affine invariant representation provided by the tensor null space operator, TNSI, we can rely on numerous methods for indexing and classification. We choose a method for dimensionality-reduction and classification based on PCNSA [3]. Notice that the term null space used in PCNSA implies that the Approximate Null Space (ANS) used for the representation of each class is formed from the minimal eigenvectors within the class, and thus minimizes the intra-class variance. However, this process is not intended to capture the null operator and is unrelated to TNSI proposed in Section 2.

First, TNSI is converted to $H = \prod_{i=1}^{n+1} W_i$ column vector Y_p which is assumed in class C_i and has Gaussian distribution as $Y|\{Y \in C_i\} \sim N(\mu_{full,i}, \Sigma_{full,i})$, where $\mu_{full,i}$ is the class conditional mean and $\Sigma_{full,i}$ is the class conditional covariance matrix. To decrease the high dimensionality, we perform Principal Component Analysis (PCA), which removes the noise-only directions and retain the directions that yield large inter-class variance. PCA takes the L leading eigenvectors of covariance matrix, Σ_{full}, of the entire data taken from all classes. The total scatter matrix, Σ_{full}, can be written $\Sigma_{full} = \Sigma_{full,w} + \Sigma_{full,b}$ where $\Sigma_{full,w}$ is within class covariance matrix and $\Sigma_{full,b}$ between class covariance matrix:

$$\Sigma_{full,w} = \frac{1}{CK} \sum_{i=1}^{C} \sum_{k=1}^{K} (Y_{i,k} - \mu_{full,i})(Y_{i,k} - \mu_{full,i})^T \ , \quad (18)$$

$$\Sigma_{full,b} = \frac{1}{C} \sum_{i=1}^{C} (\mu_{full,i} - \mu_{full})(\mu_{full,i} - \mu_{full})^T \ , \quad (19)$$

where i is for class index and k is for motion event tensor index in the class. It is assumed that there are C classes in the system and each class has K tensors.

PCA gives the L-dimensional projection matrix $(W_{PCA})_{P \times L}$. After projections, in the PCA space PCNSA finds for each class i an M_i dimensional subspace along which the class's intra-class variance is smallest. This subspace is referred to as the Approximate Null Space (ANS) denoted as N_i since the lowest eigenvalues' corresponding eigenvectors are taken. That means we choose the lowest noise variance directions as for ANS. **PCNSA Algorithm:**

1. **Obtain PCA Space:** Evaluate the total covariance matrix Σ_{full}, then apply PCA to Σ_{full} to find W_{PCA}, whose columns are the L leading eigenvectors. Project the data vectors, class means and class covariance matrices into the corresponding data vectors, class means, and class covariance matrices in the PCA space.

2. **Obtain ANS and Valid Classification Directions in ANS:** Find the approximate null space $(N_i)_{L \times M_i}$, for each class i by choosing M_i smallest eigenvalues' corresponding eigenvectors. $N_i = (e_{i,1}|e_{i,2}| \ldots |e_{i,M_i})_{L \times M_i}$. If e_i satisfies $|(\mu_i - \mu_j)^T e_i| > \delta_2 \|\mu_i - \mu_j\|$, this direction is valid and used to build valid ANS, $W_{NSA,i}$.

3. **Classification:** PCNSA finds distances from a query tensor to all classes and minimum distance to a class is chosen for classification of X.

$$d_i(X) = \|W_{NSA,i}(X - \mu_i)\| \ . \quad (20)$$

4. SIMULATION RESULTS

In order to implement and evaluate the proposed classification and retrieval system, we have used trajectories from the Australian Sign language (ASL) data set obtained from University of California at Irvine's Knowledge Discovery in Databases (UCI-KDD) archive [4]. The trajectories in the data set are obtained by registration of the hand coordinates at each successive instant of time by using a Power Glove interfaced to the system. In our simulations, we have used 40 different classes representing signing of 40 different words in the data set. Each class has 69 trajectories recorded at different instances.

Since in real life trajectories may have different lengths, we normalize the length by taking the Fourier Transform and choosing the biggest n=18 coefficients and then taking the Inverse Fourier Transform so that all the trajectories are of size 32 before invariant matrix calculations. We form the motion event tensor T by randomly selecting the trajectories from the specific class and setting J=3, P=18, K=20, namely, each tensor $T \in R^{6\times18\times20}$ contains $3 \times 20 = 60$ motion trajectories, each trajectory has 18 samples and totally there are 20 video clips. According to the definition of TNSI, applying affine transformation to the unfolding matrix $T_{(1)}$ with the dimension $2J \times PK$ indicates rotation and translation of each trajectory at all video clips to the same amount. Applying affine transformation to the unfolding matrix $T_{(3)}$ with the dimension $K \times 2PJ$ indicates rotation and translation of the same trajectory at each video clip independently.

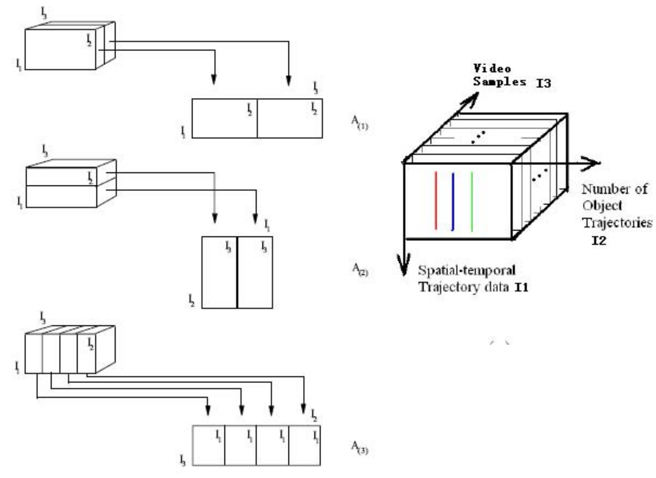

Fig. 1. Flattening a 3^{rd} order motion event tensor for multiple trajectories representation.

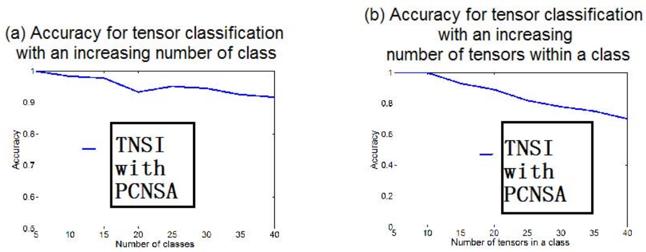

Fig. 2. Accuracy for tensor (multiple motion trajectories) classification (a) with an increasing number of classes and (b) with an increasing number of tensors within a class.

Therefore, we compute the two dimensional invariants with $\delta_2 = 10^{-4}$ as thresholds and $L = 50$ in PCNSA. Fig. 2 (a) depicts the accuracy of the proposed classification system versus the number of classes. There are 20 tensors in each class. Simulation results show that our system preserves its efficiency even for higher number of different classes. Fig. 2 (b) depicts accuracy values versus increase in the number of tensors within a class. There are 20 classes in the system. Fig. 3 shows Precision vs. Recall curves for indexing and retrieval problem by using 40 classes, each class having 20 tensors. For retrieval problems, we compute the distance of the query tensor to any other tensor using PCNSA on TNSI as $D(X_i, Y) = \|W_{NSA,i}(X_i - Y)\|$, where Y is the query tensor. This distance is then used to find α nearest tensors, where α is a user specified parameter. There are three curves in Figure 3, one is with using PCA on TNSI directly, where PCA is basically used for dimension reduction. The other two curves are applying PCNSA on single and two dimensional invariants. We illustrate in Fig. 3 that the result of using PCNSA

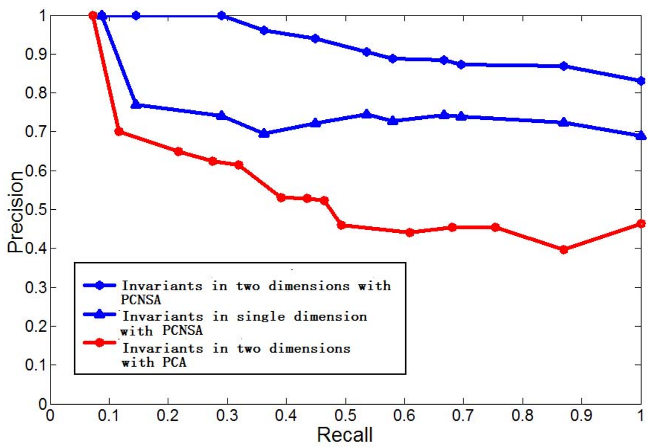

Fig. 3. Precision-Recall metric for multiple motion trajectories retrieval using PCNSA on TNSI.

on TNSI is much superior to the one using PCA on TNSI directly. Moreover, the performance of TNSI in two dimensions is much better than the single dimensional invariants.

5. CONCLUSION

We proposed the theoretical framework for tensor null space invariants as a powerful view-invariant representation for recognition and retrieval of multidimensional data sets. The proposed TNSI is perfectly multidimensional view invariant due to camera motions. We investigated the dimensions and the size of TNSI. We subsequently determined exact solutions for TNSI for arbitrary multi-dimensions of affine view invariance. Moreover, we showed that any classification algorithms can be used based on PCA representation of the TNSI operator. We further demonstrated classification and retrieval performance of multiple motion trajectories based on TNSI and PCNSA.

6. REFERENCES

[1] P.F.Stiller, C.A.Asmuth, and C.S.Wan, “A general theory of single view recognition - the affine case - with applications to indexing image databases for content based retrieval,” International Journal of Computer Vision, 1994.

[2] M.Alex, O.Vasilescu, and D.Terzopoulos, “Multilinear subspaces analysis of image ensembles,” IEEE Conference on Computer Vision and Pattern Recognition, 2003.

[3] N.Vaswani and R.Chellappa, “Principal components null space analysis for image and video classification,” *IEEE Trans. Image Processing*, 2006.

[4] S.Hettich and S.D.Bay, “The uci kdd archive,” University of California, Department of Information and Computer Science, Irvine, California, 1999.