Advances in subspace learning and its applications

Bernardo B. Gatto[1]
Federal University of Amazonas, Brazil
bernardo@icomp.ufam.edu.br

Kazuhiro Fukui
University of Tsukuba, Japan
kfukui@cs.tsukuba.ac.jp

Waldir S. S. Júnior & Eulanda M. dos Santos
Federal University of Amazonas, Brazil
{waldirjr@, emsantos@icomp.}ufam.edu.br

***Abstract*—Pattern-set matching refers to a class of problems where learning takes place through sets rather than elements. Much used in computer vision, this approach presents robustness to variations such as illumination, intrinsic parameters of the signal capture devices, and pose of the analyzed object. Inspired by applications of subspace analysis, three new collections of methods are presented in this thesis[1] summary: (1) New representations for two-dimensional sets; (2) Shallow networks for image classification; and (3) Tensor data representation by subspaces. New representations are proposed to preserve the spatial structure and maintain a fast processing time. We also introduce a technique to keep temporal structure, even using the principal component analysis, which classically does not model sequences. In shallow networks, we present two convolutional neural networks that do not require backpropagation, employing only subspaces for their convolution filters. These networks present advantages when the training time and hardware resources are scarce. Finally, to handle tensor data, such as videos, we propose methods that employ subspaces for representation in a compact and discriminative way. Our proposed work has been applied in problems other than computer vision, such as representation and classification of bioacoustics and text patterns.**

I. Introduction

The task of recognizing objects from one image has a limited capacity for recognition. For instance, single-view information may be insufficient to solve possible ambiguity due to the camera's point of view or occlusion. An image set (more generally, pattern set) is a collection of images of the same object or event. This set can be unordered, where the timestamp is not relevant or ordered. A pattern set model requires robustness to corrupt data and handle a variable set size properly without increasing computational complexity.

Subspace representation has been a common strategy to model pattern sets. A subspace eases the issues above using the geometrical structure under which images in a set are distributed. A subspace describes the set with a fixed dimension, a model with mainly two valid points. (1) Statistical robustness to input noises, i.e., perturbations such as occlusion. (2) Compactness (low dimension), even when processing many images, leads to a fixed complexity. Current challenges exist in pattern set modeling. For instance, employing Principal Component Analysis (PCA) [1] to model the pattern sets may be insufficient to represent two-dimensional patterns existing in images. PCA demands pattern vectorization, which may lead to weakening the set representation.

In this work, among other contributions, we describe a new type of subspace that can process two-dimensional image sets without damaging their two-dimensional structures. We name our model Two Dimensional Mutual Subspace Method (2D-MSM) as a mention to the Mutual Subspace Method (MSM) [2], a fundamental subspace classification algorithm. Similar to MSM, in 2D-MSM, both the input and the learnable basis vectors span subspaces and their mutual canonical angles perform their matching. Besides, we describe a variant called Hankel Mutual Subspace Method (HMSM), where the frames of an input video are arranged in a Hankel-like matrix, protecting its ordering during the extraction of its basis vectors.

Subspaces have been incorporated in shallow neural networks, concretely as parameters of Convolutional Neural Networks (CNN). We describe a shallow network based on subspaces applied in image classification problems. This new concept not only learns the network weights without using backpropagation but can work under scarce training sample conditions. The proposed network presents a discriminative space, where the extracted features provide more reliable information for classification. We also developed a convolutional neural network able to process semi-supervised data efficiently.

Many applications employ data in a tensor format, such as video and audio. Tensor data is observed in action analysis from video data, where both spatial and temporal information is present in a structured form. The spatial and temporal information can be handled independently within different representations. Inspired by the Fukunaga-Koontz Transformation (FKT) [3], we describe a formulation of FKT to handle tensor data. Our method has been applied to tensor representation to solve action learning from videos. Here, tensor data is decomposed into several subspaces, enabling tensorial learning. We also developed another solution for tensor data when only unsupervised training data is available.

General objective: Develop advances in subspace learning by introducing new representations and shallow networks. These representations reduce the complexity of solving pattern set classification and related problems. We explore different approaches to describe and classify pattern sets in diverse machine learning scenarios. **Specific objectives:** (1) Investigate variants of subspace-based methods that represent two-dimensional data. (2) Introduce shallow networks capable of learning through subspaces without employing backpropagation. (3) Propose methods for representing tensorial data.

[1]This article summarizes the main contributions of the Ph.D. Thesis entitled "Pattern-set Representations using Linear, Shallow and Tensor Subspaces".

II. Background Theory

This section describes the background theory behind subspace learning and applications on pattern set representation and classification. With video cameras being widely used, it is natural to solve a classification problem using pattern sets.

The Mutual Subspace Method (MSM) [2] is a common technique employed to represent and classify pattern sets. We define a pattern set as a collection of samples relating to a particular category. In this approach, a set of patterns is analyzed in a batch instead of separately. Matching pattern sets arises naturally in distinct circumstances, such as when the target pattern is available in a data stream. Another practical example is when the data is contained in a bag, such as the profile pictures in a social media network. The MSM's theory was developed from the observation that patterns of the same object produce a compact cluster in high-dimensional vector space. This compact cluster can be described by a subspace, which is generated by using PCA. The benefits of subspace-based methods include their high compactness ratio, low complexity, and flexibility to handle different data types.

A. Mutual Subspace Method

To represent a pattern defined by a subspace, we use the observation that a set of images is in a cluster, where orthonormal basis vectors can efficiently describe them. Let us consider an image set $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$, where x is a feature vector, possibly obtained through the vectorization of an image. Then, it is reasonable to conduct a decomposition to gather knowledge of the geometric structure of $\mathbf{X}$. The Singular Value Decomposition (SVD) [] will produce a set of eigenvectors $\mathbf{U} = \{u_1, u_2, \ldots, u_n\}$, and a set of eigenvalues $\mathbf{\Lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, where each vector in $\mathbf{U}$ represents an axis and each value in $\mathbf{\Lambda}$ describes how each axis is transformed in the space described by $\mathbf{X}$. Another useful idea is that $\mathbf{\Lambda}$ represents how much the vectors in $\mathbf{X}$ are correlated, which is a valuable guidance towards the redundant and non-redundant information in $\mathbf{X}$. The SVD of $\mathbf{X}$ is

$$\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \ . \tag{1}$$

Each column of $\mathbf{U}$ is a singular vector of $\mathbf{X}\mathbf{X}^\top$ and the main diagonal of $\mathbf{\Sigma}$ presents the singular values in descending order. The ordered nature of the singular vectors in $\mathbf{\Lambda}$ can be directly employed to reveal the importance of each eigenvector in $\mathbf{U}$. The analysis of $\mathbf{\Lambda}$ is helpful in various problems, such as dimensionality reduction, signal filtering, and feature extraction. By analyzing the influence of each eigenvector, it is possible to select a small set of eigenvectors by removing all but the top k singular values in the diagonal of $\mathbf{\Lambda}$.

B. Selecting basis vectors

The basis vectors generated by SVD may represent a set of patterns compactly. The following criteria can be utilized to obtain the compactness ratio of this transformation

$$\mu(k) \leq \sum_{i=1}^{k}(\lambda_i) / \sum_{i=1}^{n}(\lambda_i) \ . \tag{2}$$

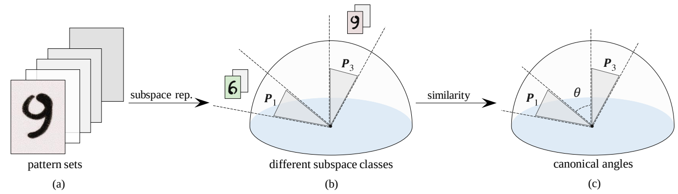

Fig. 1. Here is an effort to represent the MSM algorithm since it is difficult to draw the subspaces in a high-dimensional space. (a) An unordered set of images representing a particular character is processed (b) The subspaces are produced by extracting the basis vectors from the set of patterns. (c) The canonical angles are employed to achieve the similarity between $\mathbf{P}_1$ and $\mathbf{P}_3$.

In Eq. (2), k is the number of the selected basis vectors which will span a subspace, λ_i corresponds to the i-th eigenvalue of $\mathbf{X}\mathbf{X}^\top$. Then, $n = \text{rank}\left(\mathbf{X}\mathbf{X}^\top\right)$. It is useful to set k as small as possible to achieve a minimum number of orthonormal basis vectors, maintaining low memory requirement.

C. Computing the similarity between subspaces

A general method for estimating the similarity between subspaces is by computing the principal angles, also known as canonical angles. The canonical angles give information concerning the relative location of two subspaces in a Euclidean space. Given two subspaces, $\mathbf{P}$ and $\mathbf{Q}$ spanned by $\mathbf{U}$ and $\mathbf{V}$, the principal angles between these subspaces can be obtained by computing the eigenvalues of $\mathbf{U}^\top\mathbf{V}$ as follows:

$$\mathbf{R}\mathbf{\Sigma}\mathbf{S} = \mathbf{U}^\top\mathbf{V} \ . \tag{3}$$

The matrix $\mathbf{\Sigma}$ provides the set of eigenvalues, $\sigma_1, \sigma_2, \ldots, \sigma_k$, in its main diagonal, with $0 \leq \sigma_1 \leq \sigma_2 \leq \ldots, \leq \sigma_{k-1} \leq \sigma_k \leq 1$. Then, the canonical angles are

$$\theta_i = \cos^{-1}(\sigma_i) \ , \quad \forall \ i = 1, \ldots, k \ . \tag{4}$$

III. New Subspace Representations

Kernel Orthogonal MSM (KOMSM) has been used in many applications due to its flexibility and straightforward implementation. However, its performance is not satisfactory for more advanced systems, wherein more complicated structures should be classified. In short, KOMSM employs PCA to generate the subspaces where each two-dimensional image from a set is reshaped to a one-dimensional vector, destroying the spatial information. Similarly, KOMSM cannot handle ordered information since the SVD is ordering invariant.

To solve these issues and motivated by 2D-PCA [4] and the Hankel representation, we propose the Two-Dimensional Mutual Subspace Method (2D-MSM) and its kernelized version, Kernel Two-Dimensional Subspace (K2DS). The main difference between PCA and 2D-PCA is that 2D-PCA uses the image matrix directly, without pattern vectorization. Since MSM and KOMSM operate on the basis vectors given by PCA, replacing PCA with 2D-PCA reduces the memory cost since the basis vectors of 2D-PCA are more compact.

To extract temporal information from ordered patterns, we develop a Hankel matrix formulation for subspace representation. The image patterns can be stored in a way where the ordering of the images is preserved. We select representative samples from each gesture set to form its corresponding Hankel matrix to obtain a smaller covariance matrix.

Computational Advantage: The main difference of 2D-MSM from traditional MSM is that 2D-MSM does not require transforming image matrices into vectors. Thus, it reduces the computational complexity of constructing the subspaces and reduces the matching time. All these aspects make the proposed algorithm superior to MSM in terms of learning and testing time. More precisely, the process of extracting the basis vectors of each 2D-PCA variant determines its processing time and the dominant complexity of each algorithm.

The problem formulation of matching time-aware pattern sets is similar to the conventional pattern set matching described in Figure 1, except that the ordering of the patterns should be preserved. For instance, some gesture classes present their semantic information correlated to the pattern ordering.

Here we summarize the Hankel representation for ordered sets. A gesture that is handled as a time series of vectors can be regarded as the output of a Linear Time Invariant (LTI) system of unknown parameters. Then, given an ordered sequence of feature vectors $\mathbf{A} = \{\mathbf{A}_i\}_{i=1}^{M}$, its block-Hankel matrix is

$$\mathbf{H_A} = \left[(\mathbf{A}_1, \ldots, \mathbf{A}_{n-1})^\top, \ldots,\ (\mathbf{A}_{m+1}, \ldots, \mathbf{A}_M)^\top\right] \ , \quad (5)$$

where n is the maximal order of the system (or the number of feature vectors), M is the temporal length of the sequence, and it holds that $M = n + m - 1$.

To represent an ordered image set $\mathbf{A}$ by a subspace, we introduce the of Hankel subspace for gesture recognition. Given a Hankel matrix $\mathbf{H_A}$ from $\mathbf{A} = \{\mathbf{A}_i\}_{i=1}^{M}$, we can compute an auto-correlation Hankel matrix as: $\mathbf{C_A} = \mathbf{H_A}\mathbf{H_A}^\top$, where its eigendecomposition generates a set of eigenvectors $\mathbf{\Phi_A} = \{\phi_i\}_{i=1}^{K}$ that spans $\mathbf{P_A}$. The Eq. (3) and Eq. (4) are then employed for measuring the similarity between two Hankel subspaces. When creating a Hankel matrix, the number of images in a set and its dimension are significant computational resources. To alleviate this issue, we introduce two approaches: (1) Random sample selection, where we randomly select images from the set, preserving its original order and (2) Clustering selection, where the centroids obtained by a k-means clustering are used to represent the set.

A. Experimental Results Summary of 2D-MSM

We conducted image set matching experiments on 7 datasets: ALOI, RGB-D for object recognition, Honda/UCSD, YouTube Celebrities, PubFig83, and CMU-MoBo for face recognition and ASL Finger Spelling dataset. The classification time of the subspace methods based on 2D-PCA is about 3 times faster than the learning time and matching time of KOMSM. The classification time of 2D-MSM is about 4 times faster than the learning time and matching time of MSM, revealing that our method is more efficient than the traditional KOMSM and MSM.

B. Experimental Results Summary of Hankel subspaces

We employed Cambridge gesture for general gestures classification and Human-Computer Interaction dataset, which contains computer interface gestures. We compare the proposed Hankel subspaces with several state-of-the-art subspace-based methods, which achieved competitive accuracy, similar to discriminative methods. This indicates that the temporal information extracted by the Hankel representation is compelling, even when random samples or centroids are selected as long as the temporal order is preserved. We want to emphasize that Hankel subspaces do not employ any learning scheme different from the compared methods. This demonstrates the effectiveness of using the Hankel subspace for gesture representation.

IV. Shallow Neural Networks Based on Subspaces

Deep learning-based approaches, especially those using deep Convolutional Neural Networks (CNN), have been employed in image classification problems. Learning through deep neural networks has received significant attention due to its improvements over hand-crafted features. Despite encouraging results, the fine-tuning of deep neural network parameters is time-consuming. Many shallow networks have been proposed based on PCA, where convolutional kernels are obtained from PCA, ICA, or DCT basis vectors. For instance, PCANet (PCA network) [5] uses a CNN architecture with no pooling layers, no activation functions, and without using backpropagation. Although only PCA or Linear Discriminant Analysis (LDA) are employed, they present competitive performance compared to the state-of-the-art results.

To improve the discriminant potential of such networks, we propose a shallow network based on the Fukunaga-Koontz Transform (FKT) [6] to generate discriminative features and handle complex distributions. In our method, instead of employing PCA or LDA to learn the convolutional kernels, we use the subspace generated by FKT. Using the FKT decorrelation subspace, we build a shallow network, FKNet, that minimizes the correlation between different image classes. In FKNet, the training images are firstly compressed as subspaces to minimize their within-class distance. Besides, the decorrelation subspace based on the compressed data is more robust to outliers. The FKT kernels can reveal more discriminative information compared to related networks.

In summary, our contributions in this context are: (1) A new shallow network for image classification. Through the use of FKT, we generate a discriminative subspace projection to enhance the discriminability across the image classes. (2) A new type of convolutional kernel based on orthogonalization of subspaces. We show that the basis vectors of FKT are useful as convolutional kernels, efficiently handling supervised data, solving one of the limitations of PCANet.

Fukunaga-Koontz Network: Figure 3 shows the conceptual diagram of the proposed shallow network. FKNet processes images as follows. An input image is processed by

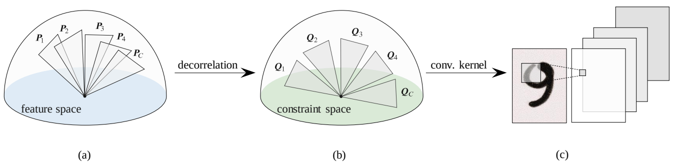

Fig. 2. The FKT and its application. (a) Image sets can be represented by $\mathbf{P}_i$ subspaces. (b) FTK is employed to decorrelate the subspaces. (c) When subspaces $\mathbf{P}_1, \ldots, \mathbf{P}_C$ represent image patches, the FKT transformation matrix can be used as a convolutional kernel.

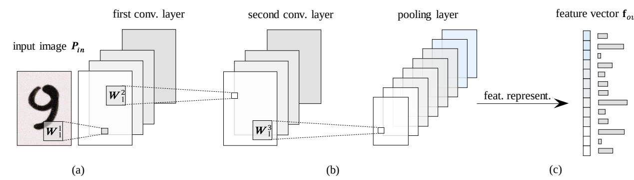

Fig. 3. The proposed network architecture: A convolutional feature extraction layer processes an input image based on FK convolutional layer, followed by another FK layer. Then, an average pooling layer is employed. Finally, binary hashing and a block-wise histogramming produce the final feature vector.

a convolutional feature extraction layer, followed by a mean-pooling or other convolutional layers. Then, binary hashing is applied to the produced features to achieve dimensionality reduction. Finally, block-wise histogramming is employed to create the final feature vector.

A. Representation by image patches

Given a dataset $\mathbf{X}$ consisting of N labeled training images of size $H \times W$, we extract patches of size $K_1 \times K_2$ from $\mathbf{X}$. This procedure is performed by taking a patch around each pixel from all N training images. Here, we denote the set of image patches as $\mathbf{P}$. Given that each image patch will have size $K_1 \times K_2$, the set $\mathbf{P}$ will contain $N_{\mathbf{P}} = HWN$ patches.

B. Computing image patches subspaces

To create subspaces, we will use the patch set $\mathbf{P} = \{p_i^j\}_{i,j=1}^{N_j,C}$, where C stands for the number of classes and N_j is the number of patches in the $j-$th class. In this C class classification problem, it is required to compute C feature matrices $\{\mathbf{A}_j\}_{j=1}^{C}$. For each feature matrix $\mathbf{A}_j$, we compute the auto-correlation matrix $\mathbf{C}_j = \mathbf{A}_j{}^\top \mathbf{A}_j$. Equipped with all C auto-correlation matrices, we can move forward to calculate the matrix $\mathbf{U}_j$ of eigenvectors which diagonalizes the auto-correlation matrix $\mathbf{C}_j$ as follows: $\mathbf{D}_j = \mathbf{U}_j{}^{-1}\mathbf{C}_j\mathbf{U}_j$ with $j = 1, \ldots, C$ and each $\mathbf{U}_j$ is a $K_1K_2 \times K_1K_2$ matrix satisfying $\mathbf{U}_j\mathbf{U}_j{}^\top = \mathbf{U}_j{}^\top\mathbf{U}_j = \mathbf{I}$. The columns of $\mathbf{U}_j$ that correspond to nonzero singular values compound a set of orthonormal basis vectors for the range of $\mathbf{C}_j$. $\mathbf{D}_j$ is the diagonal matrix of eigenvalues of $\mathbf{C}_j$. Unlike PCANet, FKNet creates a subspace for each class independently, exploiting its intrinsic characteristics in a more effective way.

C. FKT for image patches subspaces decorrelation

Once equipped with all the C image patches subspaces $\mathbf{P}_j$ and their R_j dimensions have been computed (using Eq. (2)), we can use FKT to generate the matrix $\mathbf{F}$ that can decorrelate the subspaces. Then, each set of basis vectors $\mathbf{U}_j$ spans a reference subspace $\mathbf{P}_j$. The method to generate the matrix $\mathbf{F}$ that efficiently decorrelates the C R_j-dimensional classes subspaces is explained as follows. First, we compute the total projection matrix as: $\mathbf{G} = \sum_{j=1}^{C} \mathbf{U}_j\mathbf{U}_j^\top$. The eigendecomposition of $\mathbf{G}$ produces a $K_1K_2 \times K_1K_2$ sum subspace $\mathbf{B}$ and its eigenvalues $\mathbf{\Lambda}$. By weighting $\mathbf{B}$ regarding the inverse of its eigenvalues contribution, we of obtain $\mathbf{F}$ as follows, $\mathbf{F} = \mathbf{\Lambda}^{-1/2}\ \mathbf{B}^\top$. Figure 2 illustrates the procedure to construct FTK and its application as convolutional kernels.

D. Fukunaga-Koontz convolutional kernels

After obtaining $\mathbf{P}$ and $\mathbf{F}$, we can compute the FK convolutional kernel. In our formulation, each basis vector of $\mathbf{F} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{N_\mathbf{F}}\}$ will be a convolutional kernel in the network. According to this formulation, the definition of the Fukunaga-Koontz convolutional kernel is: $\mathbf{W}_l = \text{map}_{K_1 \times K_2}(\mathbf{w}_l)$ with $l = \{1, 2, ..., L_S\}$, where the operator $\text{map}_{K_1 \times K_2}(\cdot)$ maps an input vector $y \in \mathbb{R}^{K_1K_2}$ onto a matrix $\mathbf{Y} \in \mathbb{R}^{K_1 \times K_2}$ and L_S is the number of convolutional kernels in the S-th layer.

Given an input image $\mathbf{P}_{in}$, the output image $\mathbf{Y}_l$ of a convolutional layer is obtained by $\mathbf{Y}_l = \rho(\mathbf{W}_l * \mathbf{P}_{in})$ with $l = \{1, 2, ..., L_S\}$, where $*$ refers to a convolution and $\rho(\cdot)$ is an average pooling operator, which may or may not be present in a particular layer, defined by a $B_1 \times B_2$ window.

E. Experimental Results Summary

FKNet provides competitive classification results compared to related shallow networks. To show its flexibility, FKNet was evaluated on a face verification task using the LFW dataset. FKNet was demonstrated to be competitive in this experiment, where FVF, MBSIF-OB and other shallow networks were employed as baselines. The processing time measurement by the proposed network is efficient. For instance, CNN required about 3 hours to generate a 4 convolutional layers model using the EMNIST training dataset. On the other hand, FKNet obtained a comparable model using less than 17 minutes on the same hardware, roughly one order of magnitude faster. Another benefit of using the proposed network is that the number of convolutional kernels employed is much smaller than those used by a CNN.

V. Tensor analysis based on subspaces

Tensors, which can be defined as a generalization of matrices, allow a natural representation of multi-dimensional data. For instance, video data is intuitively described by its correlated images over the time axis. Vectorization and concatenation of the video pixels may be applied to produce a practicable representation. However, the vectorization procedure may degrade the spatio-temporal relationship between pixels of video tensor data, causing information loss.

The order of a tensor is linked to its dimensions, also known as ways or modes. Tensor unfolding is a procedure

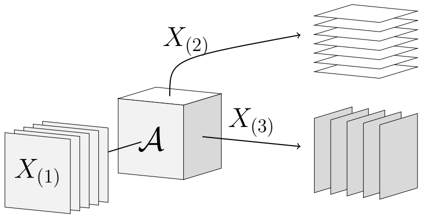

Fig. 4. The unfolding procedure of a 3-mode tensor. The unfolding of the 3-mode tensor $\mathcal{A}$ produces 3 sets of matrices $X_{(1)}$, $X_{(2)}$ and $X_{(3)}$.

that reorganizes the tensor data to permit the analysis of each mode separately, possibly revealing correlations that were not immediately observed. This tensor unfolding procedure is shown in Figure 4. The tensor unfolding maneuver facilitates the interpretability of the modes, as in medical image analysis.

Product Grassmann Manifold (PGM) efficiently represents tensor data, as in action recognition problems [7]. PGM describes subspaces as a point on the product space of n Grassmann manifolds, where each subspace corresponds to a point on one of the n Grassmann manifolds. The classification is performed based on the chordal distance [8]. Similar to MSM, PGM lacks a discriminative mechanism.

We introduce the n-mode Generalized Difference Subspace projection (n-mode GDS), extracting discriminative information from tensor data and providing suitable subspaces for tensor data classification. We employ the GDS projection, which acts as a feature extractor for MSM. Since GDS represents the difference among class subspaces, the GDS projection can increase the class subspaces' angles toward orthogonal status. Likewise in PGM, we can express tensor data as a point on a product manifold, simplifying the tensorial data representation and enabling GDS, revealing the relationship between all tensor modes in a unified design.

Our contributions are: (1) A novel tensor data representation called n-mode GDS. (2) We present a new tensor classification framework. (3) We introduce an improved version of the geodesic distance, fusing and estimating the importance of each tensor mode for an efficient classification.

A. *n-mode Generalized Difference Subspace*

Tensor data is usually represented by a set of modes (n-mode tensor) to reduce computational complexity. Given two n-mode tensors $\mathcal{A}$ and $\mathcal{B}$, we can formulate the tensor matching problem in two steps. First, we create a convenient representation, where $\mathcal{A}$ and $\mathcal{B}$ can be expressed in a compact and informative manner. Second, we establish a mechanism to produce a reliable measure of similarity between these representations, allowing the comparison of $\mathcal{A}$ and $\mathcal{B}$.

B. *Tensor Representation by Subspaces*

To simplify the tensor representation, we employ the unfolding process. We denote by $\mathbf{X} = \{X_i\}_{i=1}^{n}$ the set of unfolded images corresponding to the mode-1, mode-2 and mode-3 unfolding of $\mathcal{A}$. The same procedure is conducted on the tensor $\mathcal{B}$, resulting in $\mathbf{Y} = \{Y_i\}_{i=1}^{n}$.

We can use eigen-decomposition to derive a set of eigenvectors for each element of $\mathbf{X}$ and $\mathbf{Y}$. Then, the eigenvectors associated with the largest eigenvalues of each element of $\mathbf{X}$ and $\mathbf{Y}$ represent their elements in terms of variance maximization [1]. After selecting these eigenvectors, we obtain $\mathbf{U}_X = \{U_i\}_{i=1}^{n}$ and $\mathbf{U}_Y = \{U_i\}_{i=1}^{n}$, respectively.

Equipped with $\mathbf{U}_X$ and $\mathbf{U}_Y$, which span the n-mode subspaces $\mathbf{P} = \{P_i\}_{i=1}^{n}$ and $\mathbf{Q} = \{Q_i\}_{i=1}^{n}$, we develop a mechanism to extract discriminative information from $\mathcal{A}$ and $\mathcal{B}$ by creating a set of subspaces $\mathbf{D} = \{D_i\}_{i=1}^{n}$, whereby projecting the sets $\mathbf{P}$ and $\mathbf{Q}$ we obtain discriminative subspaces. We adopt GDS [9] since it provides a reasonable balance between robustness and computational complexity, considering that it is mainly based on eigen-decomposition. After projecting the n-mode subspaces $\mathbf{P}$ and $\mathbf{Q}$ onto $\mathbf{D}$, we obtain the sets $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$. By selecting a similarity function, we have the main tools to represent and measure the similarity between $\mathcal{A}$ and $\mathcal{B}$.

C. *Generating the n-mode GDS Projection*

In a m-class classification problem, $\mathbf{P} = \{P_{ij}\}_{i,j=1}^{n,m}$ denotes the set of all n-mode subspaces spanned by $\mathbf{U} = \{U_{ij}\}_{i,j=1}^{n,m}$. Then, we can develop the n-mode GDS projection $\mathbf{D} = \{D_i\}_{i=1}^{n}$. Since each mode subspace reflects a particular factor, it is essential to handle each one independently and compute a model that reveals hidden discriminative structures. In traditional GDS, this procedure is performed by removing the overlapping components that represent the intersection between the subspaces. In mathematical terms, the GDS projection can be described as extending the difference vector between two vectors in a multi-dimensional space.

To compute the n-mode GDS, we compute the sum of the projection matrices of each i-mode subspace as follows, $G_i = \frac{1}{m}\sum_{j=1}^{m} U_{ij}U_{ij}^{\top}$, for $1 \leq i \leq n$. Since G_i has information regarding all class subspaces in a particular mode, it is beneficial to decompose it to exploit discriminative elements. Applying eigen-decomposition to G_i, we obtain: $G_i = V_i\Sigma_i V_i^{\top}$, for $1 \leq i \leq n$, where the columns in $V_i = \{\phi_1, \phi_2, \ldots, \phi_{R_i}\}$ are the normalized eigenvectors of G_i, and Σ_i is the diagonal matrix with corresponding eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_{R_i}\}$ in descending order, where $R_i = \mathrm{rank}(G_i)$. We can define $D_i = \{\phi_{\alpha_i}, \ldots, \phi_{\beta_i}\}$, where $\alpha_i < \beta_i \leq R_i$.

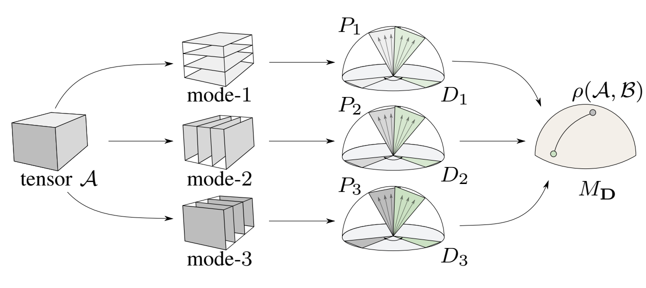

Fig. 5. Figure of the n-mode GDS projection. We unfold the $3-$mode tensor $\mathcal{A}$ and compute its subspaces. Then, we project the subspaces onto the n-mode GDS. The PGM can be exploited to represent the projected subspaces. The chordal distance $\rho(\mathcal{A}, \mathcal{B})$ determines the similarity between $\mathcal{A}$ and $\mathcal{B}$ [10].

D. Representing the n-mode Subspaces $\hat{\mathbf{P}}$ on the PGM

We introduce the product manifold to describe $\hat{\mathbf{P}}$ into a single manifold $M_{\mathbf{D}}$, which consists of the product of the projected n-mode subspaces onto the n-mode GDS (i.e., the projection of $\mathbf{P}$ onto $\mathbf{D}$). Given a set of manifolds $\mathbf{M} = \{M_i\}_{i=1}^{n}$ composed by $\hat{\mathbf{P}}$, Eq. (6) describes this procedure:

$$M_{\mathbf{D}} = M_1 \times M_2 \times \ldots \times M_n = \left(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n\right), \qquad (6)$$

where $\times$ denotes the Cartesian product, M_i is a i-mode manifold and $\hat{P}_i \in M_i$. Here, tensor data can be regarded as a point on $M_{\mathbf{D}}$, as shown in Figure 5. A benefit of employing $M_{\mathbf{D}}$ is that it allows working directly with geodesics through the use of the geodesic distance. The geodesic distance between two points is the length of the geodesic path, which is the shortest path between the points that lie on the surface of $M_{\mathbf{D}}$. Once obtained $\bar{\theta} = \{\bar{\theta}_i\}_{i=1}^{n}$ (see Eq. (3) and Eq. (4)), we can introduce the weighted geodesic distance based on the product manifold, which is defined as:

$$\rho(\mathcal{A}, \mathcal{B}) = \left(\sum_{i=1}^{n} \left(w_i \bar{\theta}_i\right)^2 \right)^{1/2}, \qquad (7)$$

where we estimate w_i by using the Fisher score [] since each mode will provide a different discriminative index.

E. Experimental Results Summary

We evaluated the proposed approach on five video datasets containing human actions and compared its results with the results achieved by other state-of-the-art approaches. The experimental results showed that the n-mode GDS outperforms conventional subspace-based methods in terms of accuracy. Moreover, the proposed n-mode GDS does not require pre-training, which is an advantage in many applications where pre-trained models are scarce.

VI. Conclusion and Future Work

We have investigated the invariant properties of pattern sets by their representations through subspaces. These invariances reflect physical properties and may be applied to represent and analyze many practical problems. Our studies have adopted a geometric framework in which the pattern sets statistical behavior is parametrized by subspaces. We handle pattern-sets as points in a metric space under this framework and analyze them using Grassmann geometry theories. The introduced methods present low computational complexity, simple implementation, and strong theoretical background.

Future work: (1) Investigate new subspace representations to express text, sound, tables, trees, and graphs, to name a few. These variants can be readily applied and employ their benefits. (2) Develop new shallow networks using the theory of Lie groups which present a simple model for continuous symmetry. (3) Introduce a deterministic neural network initialization by applying the convolutional kernels produced by FKT or GDS. (4) Propose an information fusion strategy (multimodal data) by subspaces that may handle both sound subspaces and video subspaces in a unified framework.

Acknowledgment

This work was supported by JSPS KAKENHI grant number 19K20335, the Foundation for Research Support of the State of Amazonas (FAPEAM) and the National Council for the Improvement of Higher Education (CAPES).

VII. Awards, Publications and Distinctions

Tables I and II list the works produced during the study period, composed of 5 journals and 17 conferences, denoting 22 papers. The Qualis classification is listed as well (source: qualis). Some work also received international awards.

TABLE I
Internacional Conferences

conference	qualis
ICASSP [11], [12]	**A1**
INTERSPEECH [13]	**A1**
ICIP [14]	**A1**
CVPRW [15], [16]	**A1**
ICDAR [17]	**A2**
IJCNN [18]	**A2**
ICTAI [19]	**A3**
SIBGRAPI [20]	**A3**
MLSP [21], [22]	**A4**
BRACIS [23], [24]	**A4**
MVA [25], [26], [27]	**B1**

TABLE II
Internacional Journals

journal	qualis
ESWA [10]	**A1**
ASOC [28]	**A1**
PR [29]	**A1**
NEPL [30]	**A3**
EURASIP JIVP [31]	**A4**

A. Academic Awards

1) Brazilian Conference on Intelligent Systems, BRACIS 2019. Tensor Fukunaga-Koontz Transform for Hierarchical Clustering. Award: **Best Paper Award**. URL
2) International Conference on Machine Vision Applications, MVA 2019. News2meme: An Automatic Content Generator from News Based on Word Subspaces from Text and Image. Award: **Best Poster Award**. URL
3) International Conference on Image Processing, ICIP 2017. Discriminative Canonical Correlation Analysis Network for Image Classification. Award: IEEE Signal Processing Society **Student Travel Award**, App. ID: 24775.
4) Int. Conference on Machine Vision Applications, MVA 2017. A Deep Network Model based on Subspaces: A Novel Approach for Image Classification. Award: **Best Poster Award**. URL
5) International Conference on Machine Vision Applications, MVA 2017. Enhancing Discriminability of Randomized Time Warping for Motion Recognition. Award: **Best Poster Award**. URL
6) British Machine Vision Association CVSS, BMVA CVSS 2016. A Deep Network Model based on Subspaces. Award: **Outstanding Presentation Award**. URL

B. Articles under review

1) Signal, Image and Video Process., Springer. Regularized Hankel Mutual Subspace Method for Gesture Recognition, (Qualis: **A3**)
2) Transactions on Audio, Speech and Language Processing, IEEE. Discriminative Singular Spectrum Classifier with Applications on Bioacoustic. URL, (Qualis: **A1**)

REFERENCES

[1] I. Jolliffe, "Principal component analysis," in *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.

[2] K.-i. Maeda, "From the subspace methods to the mutual subspace method," in *Computer Vision*. Springer, 2010, pp. 135–156.

[3] K. Fukunaga and W. L. Koontz, "Application of the karhunen-loeve expansion to feature selection and ordering," *IEEE Transactions on computers*, vol. 100, no. 4, pp. 311–318, 1970.

[4] 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.

[5] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[6] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.

[7] Y. M. Lui, J. R. Beveridge, and M. Kirby, "Action classification on product manifolds," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 833–839.

[8] G. H. Golub *et al.*, "Cf van loan, matrix computations," *The Johns Hopkins*, 1996.

[9] K. Fukui and A. Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2164–2177, 2015.

[10] B. B. Gatto, E. M. dos Santos, A. L. Koerich, K. Fukui, and W. S. Junior, "Tensor analysis with n-mode generalized difference subspace," *Expert Systems with Applications*, vol. 171, p. 114559, 2021.

[11] L. S. Souza, B. B. Gatto, and K. Fukui, "Classification of bioacoustic signals with tangent singular spectrum analysis," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 351–355.

[12] ——, "Grassmann singular spectrum analysis for bioacoustics classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 256–260.

[13] B. B. Gatto, E. M. dos Santos, J. G. Colonna, N. Sogi, L. S. Souza, and K. Fukui, "Discriminative singular spectrum analysis for bioacoustic classification," in *INTERSPEECH 2020*. International Speech Communication Association (ISCA), 2020.

[14] B. B. Gatto and E. M. dos Santos, "Discriminative canonical correlation analysis network for image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 4487–4491.

[15] L. S. Souza, N. Sogi, B. B. Gatto, T. Kobayashi, and K. Fukui, "An interface between grassmann manifolds and vector spaces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 846–847.

[16] N. Sogi, L. S. Souza, B. B. Gatto, and K. Fukui, "Metric learning with a-based scalar product for image-set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 850–851.

[17] B. B. Gatto, E. M. dos Santos, and K. Fukui, "Subspace-based convolutional network for handwritten character recognition," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 1044–1049.

[18] E. K. Shimomoto, L. S. Souza, B. B. Gatto, and K. Fukui, "Text classification based on word subspace with term-frequency," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[19] B. B. Gatto, W. S. da Silva, and E. M. dos Santos, "Kernel two dimensional subspace for image set classification," in *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2016, pp. 1004–1011.

[20] B. B. Gatto, E. M. dos Santos, and W. S. da Silva, "Orthogonal hankel subspaces for applications in gesture recognition," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2017, pp. 429–435.

[21] B. B. Gatto, A. Bogdanova, L. S. Souza, and E. M. dos Santos, "Hankel subspace method for efficient gesture representation," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[22] B. B. Gatto, J. G. Colonna, E. M. dos Santos, and E. F. Nakamura, "Mutual singular spectrum analysis for bioacoustics classification," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[23] B. B. Gatto and E. M. Dos Santos, "Image-set matching by two dimensional generalized mutual subspace method," in *Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2016, pp. 133–138.

[24] B. B. Gatto, M. A. Molinetti, E. M. dos Santos, and K. Fukui, "Tensor fukunaga-koontz transform for hierarchical clustering," in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2019, pp. 150–155.

[25] B. B. Gatto, L. S. de Souza, and E. M. dos Santos, "A deep network model based on subspaces: A novel approach for image classification," in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 436–439.

[26] L. S. de Souza, B. B. Gatto, and K. Fukui, "Enhancing discriminability of randomized time warping for motion recognition," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2017, pp. 77–80.

[27] E. K. Shimomoto, L. S. Souza, B. B. Gatto, and K. Fukui, "News2meme: An automatic content generator from news based on word subspaces from text and image," in *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 2019, pp. 1–6.

[28] B. B. Gatto, E. M. dos Santos, M. A. Molinetti, and K. Fukui, "Multilinear clustering via tensor fukunaga-koontz transform with fisher eigenspectrum regularization," *Applied Soft Computing*, 2021.

[29] L. S. Souza, B. B. Gatto, J.-H. Xue, and K. Fukui, "Enhanced grassmann discriminant analysis with randomized time warping for motion recognition," *Pattern Recognition*, vol. 97, p. 107028, 2020.

[30] B. B. Gatto, E. M. dos Santos, K. Fukui, W. S. Júnior, and K. V. dos Santos, "Fukunaga–koontz convolutional network with applications on character classification," *Neural Processing Letters*, vol. 52, pp. 443–465, 2020.

[31] B. B. Gatto, L. S. Souza, E. M. dos Santos, K. Fukui, W. S. Júnior, and K. V. dos Santos, "A semi-supervised convolutional neural network based on subspace representation for image classification," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, pp. 1–21, 2020.