Multimedia Systems (2020) 26:63–73
https://doi.org/10.1007/s00530-019-00629-5

SPECIAL ISSUE PAPER

Fréchet mean-based Grassmann discriminant analysis

Hongbin Yu[1] · Kaijian Xia[2] · Yizhang Jiang[1] · Pengjiang Qian[1]

Published online: 25 July 2019

Abstract
Representing image sets and videos with Grassmann manifold has become popular due to its powerful capability to extract discriminative information in machine learning research. However, existing techniques operations on Grassmann manifold are usually suffering from the problem of computational expensive, thus the application range of Grassmann manifold is limited. In this paper, we propose the Fréchet mean-based Grassmann discriminant analysis (FMGDA) algorithm to implement the videos (or image sets) data dimensionality reduction and clustering task. The data dimensionality reduction algorithm proposed by us can not only be used to reduce Grassmann data from high-dimensional data to a relative low-dimensional data, but also to maximize between-class distance and minimize within-class distance simultaneously. Fréchet mean is used to characterize the clustering center of Grassmann manifold space. We further show that the learning problem can be expressed as a trace ratio problem which can be efficiently solved. We designed a detailed experimental scheme to test the performance of our proposed algorithm, and the tests were assessed on several benchmark data sets. The experimental results indicate that our approach leads to a significant improvement over state-of-the-art methods.

Keywords Grassmann manifold · Fréchet mean · Fréchet mean-based Grassmann discriminant analysis (FMGDA)

1 Introduction

Dimensionality reduction (DR) is an important preprocessing technique in many fields of information processing such as data mining and pattern recognition. DR has been widely applied in facial recognition, feature extraction and so on. Principal component analysis, locality-preserving projection, canonical correlation analysis and linear discriminant analysis are the commonly used dimension reduction algorithms. Traditional DR algorithms are designed specifically for vector-valued data coming from a flat Euclidean space. However, in modern computer vision tasks, we often encounter the data that are sets vectors instead of data vectors. These sets vectors are usually linear representations of points in Grassmann manifold. Recent research has shown that the Grassmann manifold is an effective method to characterize videos and image sets. As such, analyzing Grassmannian points through the geometry of Euclidean spaces, such as Frobenius norm as a means to measure similarity, is inadequate. As a consequence, existing algorithms developed on the Euclidean space cannot apply directly to deal with Grassmannian points. Therefore, how to effectively and properly reduce the dimensionality of structured data becomes an urgent issue in the big data era.

A popular and geometric way to analyze Grassmannian points is using the Riemannian structure method which is induced by the principal angles. It is used to measure length of the shortest geodesic connecting the two points on the Grassmann manifold. The project metric is related to principal angles and it has been extensively studied in the literature. Harandi et al. [11] extended the classical dictionary learning approach to Grassmann manifold by embedding Grassmann points into the space of symmetric

This work was supported in part by the National Natural Science Foundation of China under Grants 61772241 and 61702225, by the Natural Science Foundation of Jiangsu Province under Grant BK20160187, by the Fundamental Research Funds for the Central Universities under Grant JUSRP51614A, by 2016 Qinglan Project of Jiangsu Province, by 2016 Six Talent Peaks Project of Jiangsu Province, and by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002.

✉ Pengjiang Qian
qianpjiang@jiangnan.edu.cn

[1] The School of Digital Media, Jiangnan University, Wuxi, Jiangsu, China

[2] China University of Mining Technology, Xuzhou, Jiangsu, China

matrices with an isometric mapping. Wang et al. [26] generalized low rank representation (LRR) model [17] on Euclidean space into the Grassmann manifold. [27] further extended the LRR model to the product Grassmann manifold. Recently, Laplacian LRR on Product Grassmann Manifolds [28] has been developed for multi-camera video surveillance.

While the Riemannian geometry induced by the projection metric has been shown promising performance in real applications, the computational cost of the resulting techniques increased drastically with the dimension of the manifold. To deal with this problem, several dimensionality reduction methods for linear subspaces have been presented. The early work includes constrained mutual subspace method (CMSM) [21] and discriminant canonical correlation (DCC) [15]. CMSM aims to find a constrained linear subspace in which the transformed sets from different classes have relatively small canonical correlations. DCC finds a linear discriminant function that maximizes the canonical correlations of within-class sets and minimizes the canonical correlations of between-class sets. However, both CMSM and DCC fail to explore the Grassmann manifold structure within image sets.

Recently, several works [3, 6, 9–12] attempt to explore the Riemannian geometry of Grassmannian manifold. Through encoding the geometry of Grassmannian manifold, existing algorithms developed in Euclidean space can be extended to deal with Grassmannian points. Hamm [9] employed the projection metric to encode the Grassmannian points by Grassmannian kernels, and developed Grassmann discriminant analysis on the kernel space. [10] further theoretically studied the relationship between projection kernel and the KL distance. Harandi [12] proposed a graph embedding based discriminant analysis approach on Grassmannian manifold which aims to simultaneously maximize discriminant power and preserve the geometrical structure of the manifold. Huang et al. [13] employed the projection metric to learn discriminant transformation on the Grassmann manifold. Wang et al. [2] extended classical LPP from the Euclidean space to Grassmann manifold.

In this paper, we propose a Fréchet Mean-based Grassmann Discriminant Analysis(FMGDA) to project high-dimensional Grassmannian points to low-dimensional Grassmann manifolds. To fulfill this goal, we introduce the Frechret mean and use the projection metric to characterize the within-class distance and between-class distance. As a consequence, classical LDA can be readily extended to non-Euclidean space. Our proposed model can be represented as a trace ratio problem. We then propose an alternating optimization approach to derive the optimal solutions. We summarize our contributions as follows:

- Based on the Fréchret mean, we introduce projection metric to characterize the within-class mean and between-class mean for Grassmannian points.
- The proposed method generalizes the classic LDA to non-Euclidean Grassmann manifolds and our optimization problem can be characterized by the trace ratio problem.

2 Backgrounds

2.1 Linear discriminant analysis

Given n samples $x_1, x_2, x_3, \ldots, x_n$ belonging to c classes, where $x_i \in \mathcal{R}^{d\times 1}$ denote the a single sample, d denote the dimension of the samples. The core idea of LDA is to seek for a linear transformation which can maximize the between-class distance and minimize the within-class distance simultaneously in the lower dimensional subspace. Denote the projection vector as a, such that the value $a^{\mathrm{T}}S_{\mathrm{b}}a$ should be maximized and the value $a^{\mathrm{T}}S_{\mathrm{w}}a$ should be minimized where S_{b} and S_{w} are named as between-class scatter matrix and with-class scatter matrix, respectively. Therefore, we have the following optimization problem (1):

$$\hat{a} = \arg\max \frac{a^{\mathrm{T}}S_{\mathrm{b}}a}{a^{\mathrm{T}}S_{\mathrm{w}}a}$$

where:

$$\begin{aligned} S_{\mathrm{b}} &= \sum_{k=1}^{c} n_k\left(\mu^{(k)}-\mu\right)\left(\mu^{(k)}-\mu\right)^{\mathrm{T}} \\ S_{\mathrm{w}} &= \sum_{k=1}^{c}\sum_{i=1}^{n_k} n_k\left(x_i^{(k)}-\mu^{(k)}\right)\left(x_i^{(k)}-\mu^{(k)}\right)^{\mathrm{T}} \end{aligned} \tag{1}$$

There $\mu^{(k)}$ is the k-th class mean vector, μ is the centroid of the total samples, $x_i^{(k)}$ is the i-th sample belonging to the k-th class and n_k denotes the number of the samples in the k-th class. With simple algebra manipulation, we can obtain the following formula:

$$\begin{aligned} a^{\mathrm{T}}S_{\mathrm{b}}a &= \sum_{k=1}^{c} n_k a^{\mathrm{T}}(\mu^{(k)}-\mu)(\mu^{(k)}-\mu)^{T}a \\ &= \sum_{k=1}^{c} n_k \|a^{\mathrm{T}}\mu^{(k)} - a^{\mathrm{T}}\mu)\|_2^2 \\ &= \sum_{k=1}^{c} n_k \delta_E(a^{\mathrm{T}}\mu^{(k)}, a^{\mathrm{T}}\mu) \end{aligned} \tag{2}$$

$$\begin{aligned} a_{\mathrm{w}}^{\mathrm{T}}a &= \sum_{k=1}^{c}\sum_{i=1}^{n_k} n_k \|a^{\mathrm{T}}(x_i^{(k)}-\mu^{(k)})\|_2^2 \\ &= \sum_{k=1}^{c}\sum_{i=1}^{n_k} n_k \|a^{\mathrm{T}}x_i^{(k)} - a^{\mathrm{T}}\mu^{(k)}\|_2^2 \\ &= \sum_{k=1}^{c}\sum_{i=1}^{n_k} n_k \delta_E(a^{\mathrm{T}}x_i^{(k)}, a^{\mathrm{T}}\mu^{(k)}) \end{aligned} \tag{3}$$

where $\delta_E(\cdot,\cdot)$ denote the Mathematical operations of the Euclidean distance between two vectors.

Optimizing (1) with respect to vector a is an easy task. Impose a constraint to the denominator of $\frac{a^T S_b a}{a^T S_w a}$ first, that is let $a^T S_w a = 1$. It is convenient to introduce the Lagrangian associated with the constrained problem, defined as:

$$L(a) = a^T S_b a - \lambda(a^T S_w a - 1) \tag{4}$$

where the value λ is known as the Lagrange multiplier. Based on the Karush–Kuhn–Tucker (KKT) conditions, the gradient of the lagrangian can be obtained:

$$\begin{aligned} &\frac{\partial L(a)}{\partial a} = 2S_b a - 2\lambda S_w a = 0 \\ \Rightarrow &S_b a = \lambda S_w a \Rightarrow S_w^{-1} S_b a = \lambda a \end{aligned} \tag{5}$$

From (5) we can learn that projected vector a is actually the eigenvector of $S_w^{-1} S_b$, the value of a can be obtained by applying the eigen-decomposition method on the matrix $S_w^{-1} S_b$, if S_w is nonsingular. There are at most $c-1$ eigenvectors corresponding to nonzero values, since the rank of S_b is bounded from above by $c-1$. Thus, the reduced dimension by LDA is at most $c-1$, where c denotes the number of object class.

2.2 Grassmann manifold

In this section, we provide a brief summary of the basic Riemannian geometry of Grassmann manifold, for more details please refer to [6, 24, 31].

A Grassmann manifold $\mathcal{G}(p, D)$ contains all of p-dimensional linear subspaces which are embedded in d dimensional Euclidean space $\mathbf{R}^D$ [25]. For $p = 0$, the Grassmann manifold becomes the Euclidean space itself. As Grassmann manifold is abstract, there are a number of ways to realize it, the most commonly used way is the thin–tall orthogonal matrix method. For a point X in the Grassmann manifold $\mathcal{G}(p, D)$ is a subspace spanned by the orthonormal columns of a $D \times p$ matrix X such that $X^T X = I_p$, where I_p is the identity matrix of size $p \times p$.

Grassmann manifold has a nice property that is it can be embedded into a space consisting of symmetric positive semi-definite matrices. More precisely, let $X \in \mathcal{G}(p, D)$, we can define the following projection embedding,

$$\Pi : \mathcal{G}(p, D) \to \mathrm{Sym}_+(D), \Pi(X) = XX^T \tag{6}$$

where $\mathrm{Sym}_+(D)$ denotes the space of $D \times D$ symmetric positive semi-definite matrices. Since $\mathrm{Sym}_+(D)$ can be understood as a Euclidean space, a natural metric for $\mathrm{Sym}_+(D)$ is the Frobenius norm. As such, we can define the following projection metric [9]:

$$\delta_P(X_1, X_2) = \frac{1}{\sqrt{2}} \|\Pi(X_1) - \Pi(X_2)\|_F^2 \tag{7}$$

where X_1 and X_2 are two Grassmann points and $\Pi(X_i) = X_i X_i^T, i = 1, 2$. As pointed in [11], $\delta_P(\cdot,\cdot)$ denote the mathematical operations which is able to approximate the true Grassmannian geodesic distance, and become one of the most popular metrics for analyzing Grassmann manifold features [3, 9, 12].

3 FMGDA

In this section, we propose FMGDA, a supervised subspace learning for Grassmann manifold that maps a high-dimensional Grassmann point to a lower dimensional Grassmann manifold.

Given a data set $(\mathcal{X}, Y)$, where $\mathcal{X} = \{X_i\}_{i=1}^n, X_i \in \mathcal{G}(p, D)$ is a Grassmann point, $Y = [y_1, y_2, \ldots, y_n]$ is the class indicator matrix. $y_i(j) = 1$ if X_i belongs to the j-th class and 0 otherwise. Our purpose is to learn a matrix $A \in \mathbf{R}^{D \times d}$ which can perform role of mapping f: $X_i \in \mathcal{G}(p, D) \to \mathcal{G}(p, d)$.

The mapping can be defined as:

$$f(X_i, A) = A^T X_i \tag{8}$$

With this mapping f, the original high-dimensional Grassmann manifold can be transformed into a lower dimensional Grassmann manifold. It is clear that $A^T X_i$ is not a valid Grassmann point, because it is impossible to guarantee that parameter A is an orthogonal matrix. To solve this problem, we temporarily employ the orthonormal components of $A^T X_i$ denoted by $A^T X_i'$ to represent an orthonormal basis matrix of the transformed projection matrix. For convenience, we rewrite $f(X_i, A) = A^T X_i'$ to make sure that $f(X_i, A)$ is a qualified Grassmann point. The approach used to get the valid Grassmann point $A^T X_i'$ from $A^T X_i$ will be presented in Sect. 4.

Suppose $X^{(k)} = [X_1^{(k)}, X_2^{(k)}, \ldots, X_{n_k}^{(k)}]$ be a set of Grassmann points from the k-th class. Recall the definition of the classical LDA algorithm, we require the mean to capture the discriminant information for classification. However, traditional Euclidean mean is not a valid Grassmann point, that cannot perform the classification task in the Grassmann manifold space. Therefore, special care must be taken into account to compute the subspace mean for Grassmann manifold. Fortunately, subspace mean on the Grassmann manifold has been studied [8, 14, 18], of which Fréchet mean is the mostly used method to characterize Grassmann manifold subspace mean.

Definition 1 The Fréchet mean M^* for a set of points $\{X_i\}_{i=1}^n, X_i \in \mathcal{G}(p, D)$ is the local minimizer of the cost function

$$M^* = \arg\min_M \sum_{i=1}^n \delta_P(X_i, M) \tag{9}$$

The above definition shows that the subspace mean depends heavily on the metric. For example, assuming that all of points come from Euclidean space with the Euclidean metric, which Fréchet mean has a closed form solution. The Fréchet mean is nothing but the traditional mean. Unfortunately, it is usually impossible to get a closed form solution for M^* using the Riemannian metric and the first-order gradient descent method [1] is commonly employed to find the optimal solution. However, for those data points of Grassmann manifold with the projection metric, we have an analytic solution for the Fréchet mean which is characterized in the following lemma.

Lemma 1 *The Fréchet mean M^* for a set of points $\{X_i\}_{i=1}^n, X_i \in \mathcal{G}(p,D)$ is the p largest eigenvectors of $\sum_{i=1}^n X_i X_i^{\mathrm{T}}$.*

Let $M^{(k)}$ be the class Fréchet mean of the k-th samples $X^{(k)}$ and M be the total Fréchet mean of $\mathcal{X}$. Similar to the classical LDA, the within-class distance and between-class distance in the transformed low-dimensional Grassmann manifold are defined as:

$$\begin{aligned} d_{\mathrm{w}} &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} n_k \delta_P(f(X_i^{(k)},A), f(M^{(k)},A)) \\ d_{\mathrm{b}} &= \sum_{k=1}^{K} n_k \delta_P(f(M^{(k)},A), f(M,A)) \end{aligned} \tag{10}$$

where d_{w} and d_{b} denote the within-class distance and between-class distance in Grassmann manifold, respectively. Denote $f(X_i^{(k)},A) = A^{\mathrm{T}}X_i', f(M,A) = A^{\mathrm{T}}M'$ and $f(M^{(k)},A) = A^{\mathrm{T}}M'^{(k)}$. Note that $A^{\mathrm{T}}M'$ and $A^{\mathrm{T}}M'^{(k)}$ denote the orthonormal components of $A^{\mathrm{T}}M$ and $A^{\mathrm{T}}M^{(k)}$ respectively, which has been mentioned above. By the definition of $\delta_P(\cdot,\cdot)$, we can explicitly rewrite d_{w} and d_{b} as follows,

$$\begin{aligned} d_{\mathrm{w}} &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} n_k \|A^{\mathrm{T}}X_i'^{(k)}X_i'^{(k)T}A - A^{\mathrm{T}}M'^{(k)}M'^{(k)T}A\|_F^2 \\ d_{\mathrm{b}} &= \sum_{k=1}^{K} n_k \|A^{\mathrm{T}}M'^{(k)}M'^{(k)\mathrm{T}}A - A^{\mathrm{T}}M'M'^{T}A\|_F^2 \end{aligned} \tag{11}$$

It should be noted that $A^{\mathrm{T}}M'^{(k)}$ cannot be used as the representative of Fréchet mean in the lower dimensional manifold mapped from $X^{(k)}$. As the method used to get Fréchet mean from $A^{\mathrm{T}}M'^{(k)}$ will be given in the Sect. 4. *Follow the idea of LDA, we now present our FMGDA algorithm, the objective function of which is defined by*

$$\max_A \frac{d_{\mathrm{b}}}{d_{\mathrm{w}}} \tag{12}$$

Similar to the classical LDA, (12) *aims to seek for mapping matrix $A \in \mathbf{R}^{D\times d}$, map a high-dimensional Grassmann point $X_i \in \mathcal{G}(p,D)$ to a relative low-dimensional Grassmann manifold $\mathcal{G}(p,d)$ and maximize the between-class distance and minimize the within-class distance of the Grassmann manifold data, simultaneously.*

4 Iterative optimization

Solving problem (12) involves optimizing four variables, including $A, X_i'^{(k)}, M'$ and $M'^{(k)}$, which is hard to get a closed form solution. In the following, we propose an iterative optimization method that optimizing one variables at a time and fixing the rest, repeating for a certain number of iterations.

We adopt the method proposed in [13, 25] to obtain the orthonormal components of $A^{\mathrm{T}}X_i'^{(k)}, A^{\mathrm{T}}M'$ and $A^{\mathrm{T}}M'^{(k)}$. Specifically, let $A^{\mathrm{T}}X_i^{(k)} = Q_{x_i}R_{x_i}$ be the QR decomposition, where $Q_{x_i} \in \mathbf{R}^{d\times p}$ is an orthogonal matrix and $R_{x_i} \in \mathbf{R}^{p\times p}$ is a nonsingular upper-triangular matrix. It is easy to show that $Q_{x_i} = A^{\mathrm{T}}X_i'^{(k)}$, where $X_i'^{(k)} = X_i^{(k)}R_{x_i}^{-1}$. Note that Q_{x_i} and $A^{\mathrm{T}}X_i^{(k)}$ represent the same subspace and Q_{x_i} is orthonormal. Similar normalization method can be applied to $A^{\mathrm{T}}M$ and $A^{\mathrm{T}}M^{(k)}$ to get M' and $M'^{(k)}$, respectively.

For convenience, denote $B^{(k)} = M'^{(k)}M'^{(k)\mathrm{T}} - M'M'^{T}$, we can reformulate d_{b} as following:

$$\begin{aligned} d_{\mathrm{b}} &= \sum_{k=1}^{K} n_k \|A^{\mathrm{T}}M'^{(k)}M'^{(k)\mathrm{T}}A - A^{\mathrm{T}}M'M'^{T}A\|_F^2 \\ &= \sum_{k=1}^{K} n_k \|A^{\mathrm{T}}B^{(k)}A\|_F^2 \\ &= \sum_{k=1}^{K} n_k tr(A^{\mathrm{T}}B^{(k)}AA^{\mathrm{T}}B^{(k)}A) \end{aligned} \tag{13}$$

Similarly, denote $Q_{ik} = X_i'^{(k)}X_i'^{(k)\mathrm{T}} - M'^{(k)}M'^{(k)\mathrm{T}}$, we can reformulate d_{w} as

$$\begin{aligned} d_{\mathrm{w}} &= \|A^{\mathrm{T}}X_i'^{(k)}X_i'^{(k)\mathrm{T}}A - A^{\mathrm{T}}M'^{(k)}M'^{(k)\mathrm{T}}A\|_F^2 \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} n_k \|A^{\mathrm{T}}Q_{ik}A\|_F^2 \\ &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} n_k tr(A^{\mathrm{T}}Q_{ik}AA^{\mathrm{T}}Q_{ik}A) \end{aligned} \tag{14}$$

Base on the above formula, we can define the objective function g_t in the t-th iteration by using the last step $A^{(t-1)}$ as follows

$$g_t = \frac{tr(A^{\mathrm{T}}\widetilde{B}^{(t-1)}A)}{tr(A^{\mathrm{T}}\widetilde{Q}^{(t-1)}A)} \tag{15}$$

Table 1 FMGDA Algorithm

Input: Grassmann points $\{X\}_{i=1}^{n}, X_i \in \mathcal{G}_{p,D}$
Output: The mapping matrix $A \in \mathbf{R}^{D\times d}$
1:Initialize: Set the parameter $A^{(0)}$
2:Compute the class and total Fréchet mean $M^{(k)}, M$ respectively according to Lemma 1
3:**While** not converged **do**
4: Normalized $X_i^{(k)}, M, M^{(k)}$ according to **QR** decomposition
5: Compute $B^{(k)}$ and Q_{ik}
6: Solve the trace ration problem (16)
7:**end while**

where

$$\widetilde{B}^{(t-1)} = \sum_{k=1}^{K} n_k B^{(k)} A^{(t-1)} A^{(t-1)^{\mathrm{T}}} B^{(k)}$$

$$\widetilde{Q}^{(t-1)} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} n_k Q_{ik} A^{(t-1)} A^{(t-1)^{\mathrm{T}}} Q_{ik}$$

Clearly, the solution of t-th iteration can be obtained by optimizing the following objective function:

$$\max_{A} \frac{tr(A^{\mathrm{T}}\widetilde{B}^{(t-1)}A)}{tr(A^{\mathrm{T}}\widetilde{Q}^{(t-1)}A)} \tag{16}$$

(16) is a trace ratio optimization problem and it has been extensively studied in [19, 20, 29]. In this paper, we adopt the method proposed in [19] to solve the optimization problem (16). The whole procedure of FMGDA is summarized in Table 1.

5 Experiments

In this section, a series of experiments are conducted on several benchmark data sets to evaluate the performance of our proposed algorithm. Specifically, three video data sets including Highway traffic data set [4], UCF sports data set [22, 23] and Ballet data set [30] are employed to carry out the classification task and three other image data set including CALTECH-101 [7] and CIFAR10 [16] are used to perform the image recognition task in our designed experiments. The performance comparison experiments are run for 5 rounds and 10%,20%,40%,60%,80% of all the data are used as the training data respectively, each round are repeated ten times, averaging accuracy result is reported.

5.1 Experimental setup

We first show how to represent image sets and videos on the Grassmann manifold. For each image set or video clip, we use the singular value decomposition (SVD) to get the basis of the matrix which consists of raw features of image sets or videos. More precisely, let$\{\mathbf{X}_i\}_{i=1}^{M}$ be an image set, where $\mathbf{X}_i$ is a gray scale image with dimension $m \times n$ and M is the number of images. By vectorizing all the images and stacking them along the column, we get a matrix $\mathbf{\Upsilon} = [\mathrm{vec}(\mathbf{X}_1), \mathrm{vec}(\mathbf{X}_2), \ldots, \mathrm{vec}(\mathbf{X}_M)]$ which can be factorized as $\mathbf{\Upsilon} = \mathbf{U\Sigma V}^{\mathrm{T}}$ via SVD. We then choose the first p columns of $\mathbf{U} \in \mathcal{G}(p, m \times n)$ as the Grassmannian point to represent the image set $\{\mathbf{X}_i\}_{i=1}^{M}$.

Five algorithms are utilized in our experiments:

- KNN: k nearest neighbor classifier using projection metric for classifying Grassmann points without dimensionality reduction.
- PML [13]: Projection metric learning on the Grassmann manifold which can be seen an extension of Fisher LDA-like framework on the Grassmann manifold.
- GGDA [12]: Graph embedding discriminant analysis on Grassmannian manifolds.
- GLPP [25]: Locality Preserving Projections for Grassmann manifold.
- FMGDA: The proposed algorithm in this paper.

We first use PML, GGDA , GLPP and FMGDA project the high-dimensional Grassmann points to a lower dimensional manifold, then KNN with $k = 1$ is employed to classify Grassmann points.

5.2 Classification on Highway traffic data set

This set contains in total 253 video sequences captured under various weather conditions, such as sunny, cloudy and rainy. There are 253 video sequences which are labeled with three labels: light, media and heavy. Specifically, there are 44 clips of heavy level, 45 clips of medium level and 164 clips of light level. Each sequences contains 42 to 52 frames. The video sequences are converted to gray image and each image is normalized to size 24×24. Examples of the high traffic data set are shown in Fig. 1.

We regard each video sequence as an image set to construct a point on the Grassmann manifold. Specifically, we choose the first 42 frames for each sequence to construct the image set and use SVD to get the subspace representation. The subspace dimension for each image set is set to 15, hence we get a Grassmann point $X \in R^{576\times 15}$ which is the input data for the above algorithms. We randomly select p percent image sets for training and the rest are used for test. This process is repeated ten times and the averaged classification accuracy is reported.

Table 2 shows the classification accuracy on Highway traffic data set. It can be observed that learning lower dimensional Grassmann points can significantly improve the classification accuracy. Only a small number of data are used in

Fig. 1 Samples from the highway traffic data set

Table 2 Classification results on highway traffic data set

Method	10%	20%	40%	50%	60%	80%
KNN	0.7500	0.7783	0.7763	0.7778	0.7255	0.8233
GGDA	0.6798	0.6897	0.6842	0.6905	0.6373	0.6471
PML	0.6903	0.7685	**0.7992**	**0.8123**	0.7743	0.8235
GLPP	0.7021	0.7236	0.7397	0.7216	0.7275	0.7445
FMGDA	**0.7632**	**0.7790**	0.7959	0.8016	**0.7853**	**0.8453**

The best results for each experiment are indicated in bold

the training procedure, all of the algorithms achieved a considerable performance. In addition, FMGDA achieves most of the best performance among the four compared algorithms, only slightly lower than the PML when using 40% and 50% data for training. From the Table 2 we observed that when using 60% samples as training data, the classification accuracy decreases, which exists in all of the four algorithms. We guess that curious more training samples will introduce more noise, leading to the overfitting problem of the Grassmann manifold, thus reducing its classification ability. Whatever, the result can demonstrate that the proposed Fréchet mean-based discriminant analysis can capture more discrimination than the other methods.

5.3 Classification on UCF sports data set

UCF Sports data set [22, 23] consists of a set of actions collected from various sports which are typically featured on broadcast television channels such as the BBC and ESPN. The video sequences were obtained from a wide range of stock footage websites including BBC Motion gallery and GettyImages. The data set includes a total of 150 sequences with the resolution of 720×480. The collection represents a natural pool of actions featured in a wide range of scenes and viewpoints. By releasing the data set we hope to encourage further research into this class of action recognition in unconstrained environments. Since its introduction, the data set has been used for numerous applications such as: action recognition, action localization, and saliency detection. Examples of the UCF sports data set are shown in Fig. 2.

Experimental results are depicted in Table 3. It is interesting to note that with relatively small number of training samples ,all of the four algorithms have very low recognition accuracy rate, however, the performance improved with the number of training samples increase. More specifically, FMGDA achieves five leading performances in all of

Fig. 2 Samples from the UCF sports data set

Table 3 Classification results on UCF sports data set

Method	10%	20%	40%	50%	60%	80%
KNN	0.4055	0.4109	0.4222	0.4438	0.4667	0.5333
GGDA	0.4151	0.4217	0.4280	0.4623	**0.5002**	0.5659
PML	0.3893	0.3933	0.4037	0.4267	0.4333	0.5778
GLPP	**0.4231**	0.4155	0.4129	0.4050	0.4751	0.5325
FMGDA	0.4189	**0.4353**	**0.4538**	**0.4665**	0.4889	**0.6111**

The best results for each experiment are indicated in bold

the six comparisons, the recognition accuracy can achieve 61% when the training samples account for 80% of the total samples.

5.4 Classification on ballet data set

The ballet data set consists of 44 videos of 8 actions collected from an instructional ballet DVD [30]. This data set is very challenging due to the significant intra-class variations in terms of speed, spatial and temporal scale, clothing and movement. Examples of ballet data set are shown in Fig. 3.

Following [11], we extract 2400 images sets by grouping every 6 frames with the same action into one image set. We use Histogram of Oriented Gradient(HOG) [5] to represent each frame and the image set is represented by a subspace of order 6. As previous done, this data set is partitioned into the training set and test set according to the specified ratio. This process is repeated ten times and the averaged accuracy is reported.

Table 4 shows the classification results. It can be observed that GGDA and PML do not work well on this data set, KNN and GLPP are unsupervised algorithms, they are unable to exceed the semi-supervised algorithms . In contrast, employing 80% of samples as training set, our FMGDA gains the best performance on this data set. This might be due to the fact that the Fréchet mean based learning algorithm can capture more discriminant information.

5.5 Recognition on Caltech-101 data set

The Caltech101 data set [7] contains 9,144 images in 102 categories (one background), about 40 to 800 images per category, most categories have about 50 images. From Caltech-101 data set we choose those categories which have more than 80 images to perform the recognition task, there are 26 categories used in this designed experiment. Examples of Caltech-101 data set have been show in Fig. 4.

In this experiments we first use the matlab function imresize function to adjust the image into the same size 300×400, then we extract the hog features of each image, the principal component analysis (PCA) is finally adopted to reduce the dimension into 512. 10%, 20%, 40%, 60% and 80% data are randomly chosen to used as the training data, the rest are the testing data, comparison results are given in the following Table 5.

Table 5 shows that our proposed FMGDA algorithms achieve the best performance among the four algorithms and GGDA have the lowest recognition accuracy.

5.6 Recognition on Cifar-10 imagesets

CIFAR-10 is an established computer-vision data set used for object recognition. It is a subset of the 80 million tiny images data set and consists of 60000 32×32 color images containing

Fig. 3 Samples from the Ballet data set

Table 4 Classification results on Ballet sports data set

Method	10%	20%	40%	50%	60%	80%
KNN	0.5000	**0.6719**	0.8333	0.9062	0.9215	0.9375
GGDA	0.3889	0.5625	0.4792	0.6253	0.5625	0.7530
PML	0.4536	0.6260	0.7240	0.8021	0.7891	0.8282
GLPP	0.4836	0.6065	0.7833	0.8080	0.8325	0.8687
FMGDA	**0.5278**	0.6255	**0.8753**	**0.9213**	**0.9225**	**0.9522**

The best results for each experiment are indicated in bold

one of 10 object classes, with 6000 images per class [16]. The samples of the Cifar-10 data set is presented in Fig. 5.

As the cifar-10 data sets, we use the image original pixels and extract the HOG feature to implement the recognition task respectively. The comparison results are given in Table 6. From the Table we can know that our proposed FMGDA have the best performance when using the original pixels as the feature, however, PML perform better than our proposed FMGDA as well as the other two algorithms when we adopt the HOG feature. In this experiment, the PML does the best recognition job, our proposed FMGDA's recognition accuracy lower 7% than the PML, however, the PML computation cost is ten times more than FMGDA. In the future we will go on the research of the FMGDA and improve its performance.

5.7 Computation complexity analysis

To further validate our proposed FMGDA, we compare the computation complexity of our algorithm with its competitors in this section. Optimizing the FMGDA objective function involving a serious mathematical operations: From step 2 in the Table 1: FMGDA Algorithms, we need to compute the total Fréchet mean M^k and M according to Lemma 1 before the loop iterations procedure. Step 2 mainly demands some matrix multiplication and eigenvalue decomposition operations whose computation complexity are $\mathcal{O}(D^2pn)$ and $\mathcal{O}(D^3)$, respectively. At the step of 4 in loop iterations , normalize $X_i^{(k)}$,M,$M^{(k)}$ according to QR decomposition involving some matrix multiplications, inverse of the matrix and QR decomposition operations, which demands $\mathcal{O}(ndDp) + \mathcal{O}(nDp^2)$,$\mathcal{O}(p^3)$ and $\mathcal{O}(nD^3)$. Compute Q_{ik} and B^k demand the computation of $\mathcal{O}(D^2pn)$ in the step 5. Solve the trace ration problem(15) in step 6 demands some matrix multiplications as well as the Singular Value Decomposition (SVD) operations which demands $\mathcal{O}(pDdn) + \mathcal{O}(np^2d) + \mathcal{O}(np^2d)$ and $\mathcal{O}(d^3)$.

For GGDA algorithm also involve a Singular Value Decomposition (SVD), which has the computation

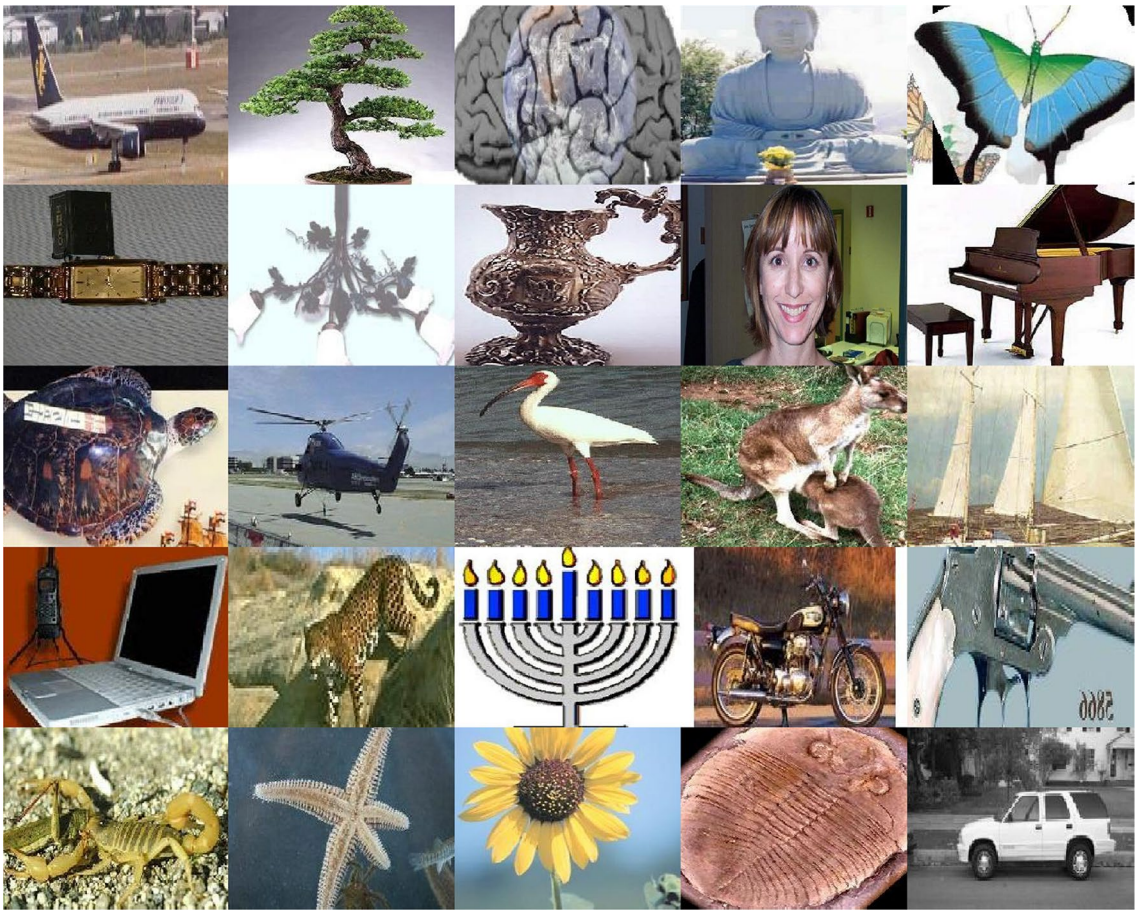

Fig. 4 Samples from the Caltech-101 data sets

Table 5 Recognition results on Caltech-101 data sets

Method	10%	20%	40%	50%	60%	80%
KNN	0.6068	0.6345	0.7362	0.7385	0.7212	0.7115
GGDA	0.3632	0.4432	0.4423	0.4538	0.4519	0.4808
PML	0.6111	0.6792	0.7254	0.7115	0.7352	0.7308
GLPP	0.5993	0.6580	0.7430	0.7466	0.7550	0.7471
FMGDA	**0.6255**	**0.6827**	**0.7596**	**0.7508**	**0.7933**	**0.7756**

The best results for each experiment are indicated in bold

complexity of $\mathcal{O}(s^3)$ for a square matrix of size $s \times s$. Solving the generalized eigenvector problem demands for $\mathcal{O}(n^3)$ operations. Computing $k^{[cc]}$ and $k^{[proj]}$ in [25] demands for $\mathcal{O}(\frac{n(n-1)}{2}p^2D + p^3)$ and $\mathcal{O}(\frac{n(n-1)}{2}p^2D)$, respectively. Considering that $p \ll D$ and $n \ll D$, the computational complexity of the proposed algorithm is hence $\mathcal{O}(p^2Dn^2)$.

Similar to GGDA and FMGDA algorithms, PML and GLPP algorithms both adopt the Riemannian Conjugate Gradient(RCG) algorithm to optimize the trace ration function and QR decomposition algorithms as well as some matrix multiplication operations. GLPP also propose SVD algorithm as an alternative algorithm to optimize trace ration function.

From the analysis mentioned above, we can see that optimize all the PML, GLPP, GGDA and FMGDA algorithms adopt the same optimization strategy such as QR, SVD as well as the RCG algorithms, that is to say the four algorithms have the same computation complexity except some general matrix multiplication operations. But RCG algorithm is more complexity than the SVD, so those algorithm who adopt the RCG to optimize the trace ration algorithm is much slower than its competitors, for example the PML algorithm.

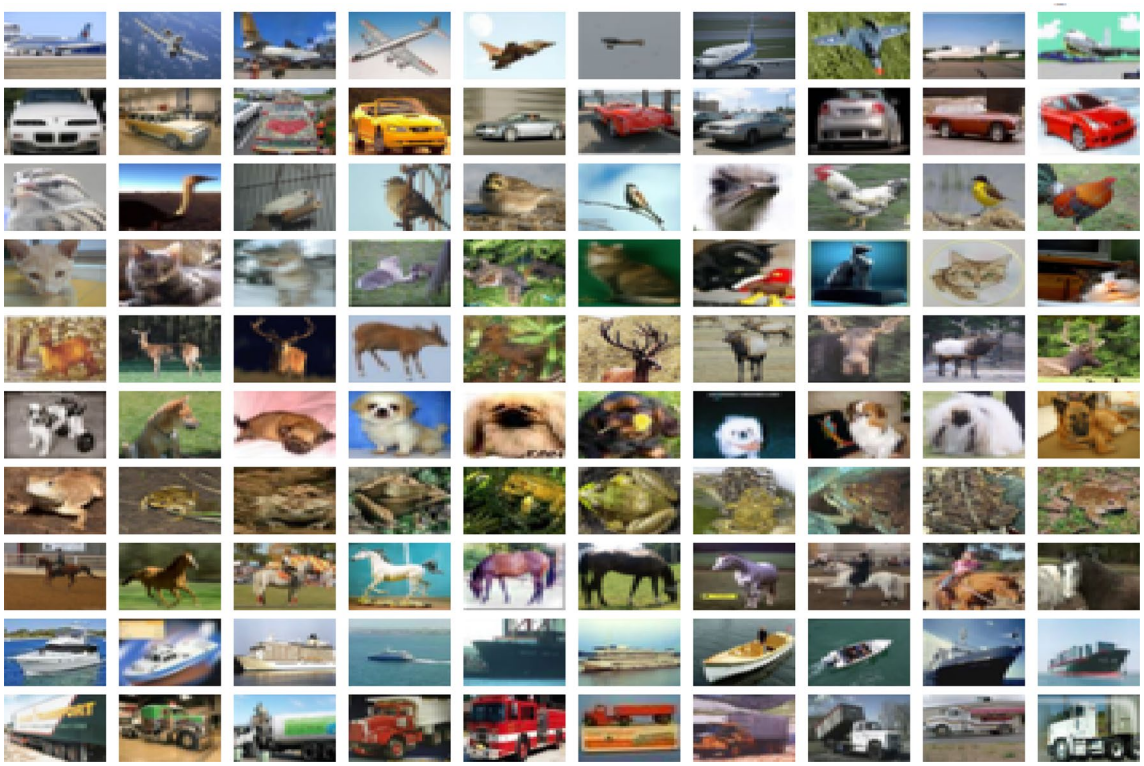

Fig. 5 Samples from the Cifar-10 data set

Table 6 Recognition results on Cifar-10 data set

Method	10%	20%	40%	50%	60%	80%
KNN	0.5081	0.5026	0.5042	0.5150	0.5135	0.5128
GGDA	0.5102	0.5188	0.5233	0.5336	0.5320	0.5389
PML	0.5398	0.5381	0.5478	0.5476	0.5559	0.5485
GLPP	0.5251	0.5389	0.5138	0.5084	0.5556	0.5322
FMGDA	**0.5508**	**0.5540**	**0.5605**	**0.5606**	**0.5573**	**0.5612**
KNN+HOG	0.7689	0.7925	0.8267	0.8460	0.7975	0.8350
GGDA+HOG	0.7250	0.7768	0.8310	0.8007	0.7967	0.8106
PML+HOG	**0.8433**	**0.8525**	**0.8690**	**0.9033**	**0.9200**	**0.9136**
GLPP+HOG	0.7655	0.7792	0.7960	0.7818	0.8001	0.8489
FMGDA+HOG	0.7998	0.8108	0.8175	0.8395	0.8865	0.8719

The best results for each experiment are indicated in bold

6 Conclusion

In this paper, we proposed a Fréchet mean-based discriminant analysis to project high-dimensional Grassmann points to a low-dimensional Grassmann manifolds. To fulfill this goal, we defined the class Fréchet mean and total Fréchet mean to characterize the center of Grassmann points. As a consequence, classical LDA can be readily extended to non-Euclidean space. Our proposed model can be represented as a trace ratio problem. We then propose an alternating optimization approach to derive the optimal solutions. Experimental results on benchmark data sets have demonstrated that our approach can lead to significant improvement over state-of-the-art methods.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Riemannian geometry of grassmann manifolds with a view on algorithmic computation. Acta Appl. Math. **80**(2), 199–220 (2004)
2. Yongli Hu, Boyue Wang, J.G.Y.S.H.C.M.A.B.Y.: Locality preserving projections for grassmann manifold. In: Proceedings

of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 2893–2900 (2017). https://doi.org/10.24963/ijcai.2017/403
3. Centingu, H.E., Vidal, R.: Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1896–1902. IEEE (2009)
4. Chan, A.B., Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures. IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 909–926 (2008)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893. IEEE (2005)
6. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998)
7. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. **106**(1), 59–70 (2007)
8. Fletcher, P.T., Venkatasubramanian, S., Joshi, S.: The geometric median on riemannian manifolds with application to robust atlas estimation. NeuroImage **45**(1), S143–S152 (2009)
9. Hamm, J., Lee, D.D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: Proceedings of the 25th International Conference on Machine Learning (ICML), pp. 376–383. ACM (2008)
10. Hamm, J., Lee, D.D.: Extended grassmann kernels for subspace-based learning. In: Advances in neural information processing systems (NIPS), pp. 601–608. Curran Associates (2009)
11. Harandi, M., Sanderson, C., Shen, C., Lovell, B.C.: Dictionary learning and sparse coding on grassmann manifolds: an extrinsic solution. In: Proceedings of the IEEE International Conference on Computer Vision (CVPR), pp. 3120–3127. IEEE (2013)
12. Harandi, M.T., Sanderson, C., Shirazi, S., Lovell, B.C.: Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2705–2712. IEEE (2011)
13. Huang, Z., Wang, R., Shan, S., Chen, X.: Projection metric learning on grassmann manifold with application to video based face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 140–149. IEEE (2015)
14. Karcher, H.: Riemannian center of mass and mollifier smoothing. Commun. Pure Appl. Math. **30**(5), 509–541 (1977)
15. Kim, T.K., Kittler, J., Cipolla, R.: Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1005–1018 (2007)
16. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images. Technical Report. University of Toronto. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf (2009)
17. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
18. Marrinan, T., Ross Beveridge, J., Draper, B., Kirby, M., Peterson, C.: Finding the subspace mean or median to fit your need. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1082–1089 (2014)
19. Ngo, T.T., Bellalij, M., Saad, Y.: The trace ratio optimization problem for dimensionality reduction. SIAM J. Matrix Anal. Appl. **31**(5), 2950–2971 (2010)
20. Nie, F., Xiang, S., Jia, Y., Zhang, C., Yan, S.: Trace ratio criterion for feature selection. AAAI **2**, 671–676 (2008)
21. Nishiyama, M., Yamaguchi, O., Fukui, K.: Face recognition with the multiple constrained mutual subspace method. In: International Conference on Audio-and Video-Based Biometric Person Authentication, pp. 71–80. Springer, New York (2005)
22. Rodriguez, M.D., Ahmed, J., Shah, M.: Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8. IEEE (2008)
23. Soomro, K., Zamir, A.R.: Action recognition in realistic sports videos. In: Computer Vision in Sports, pp. 181–208. Springer, New York (2014)
24. Srivastava, A., Klassen, E.: Bayesian and geometric subspace tracking. Adv. Appl. Probab. **36**(1), 43–56 (2004)
25. Wang, B., Hu, Y., Gao, J., Sun, Y., Chen, H., Ali, M., Yin, B.: Locality preserving projections for grassmann manifold. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 2893–2900. AAAI Press, San Francisco (2017)
26. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Low rank representation on grassmann manifolds. In: Asian Conference on Computer Vision, pp. 81–96. Springer, New York (2014)
27. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Product grassmann manifold representation and its LRR models. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp. 2122–2129. AAAI Press (2016)
28. Wang, B., Hu, Y., Gao, J., Sun, Y., Yin, B.: Laplacian LRR on product grassmann manifolds for human activity clustering in multicamera video surveillance. IEEE Trans. Circ. Syst. Video Technol. **27**(3), 554–566 (2017)
29. Wang, H., Yan, S., Xu, D., Tang, X., Huang, T.: Trace ratio vs. ratio trace for dimensionality reduction. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 1–8 (2007)
30. Wang, Y., Mori, G.: Human action recognition by semilatent topic models. IEEE Trans. Pattern Anal. Mach. Intell. **31**(10), 1762–1774 (2009)
31. Wong, Y.C.: Differential geometry of grassmann manifolds. Proc. Natl. Acad. Sci. **57**(3), 589–594 (1967)