Graph Embedding Discriminant Analysis on Grassmannian Manifolds for Improved Image Set Matching

Mehrtash T. Harandi, Conrad Sanderson, Sareh Shirazi, Brian C. Lovell

NICTA, PO Box 6020, St Lucia, QLD 4067, Australia
The University of Queensland, School of ITEE, QLD 4072, Australia

Abstract

A convenient way of dealing with image sets is to represent them as points on Grassmannian manifolds. While several recent studies explored the applicability of discriminant analysis on such manifolds, the conventional formalism of discriminant analysis suffers from not considering the local structure of the data. We propose a discriminant analysis approach on Grassmannian manifolds, based on a graph-embedding framework. We show that by introducing within-class and between-class similarity graphs to characterise intra-class compactness and inter-class separability, the geometrical structure of data can be exploited. Experiments on several image datasets (PIE, BANCA, MoBo, ETH-80) show that the proposed algorithm obtains considerable improvements in discrimination accuracy, in comparison to three recent methods: Grassmann Discriminant Analysis (GDA), Kernel GDA, and the kernel version of Affine Hull Image Set Distance. We further propose a Grassmannian kernel, based on canonical correlation between subspaces, which can increase discrimination accuracy when used in combination with previous Grassmannian kernels.

1. Introduction

In contrast to object recognition approaches based on considering one image at a time, there has been a recent surge of interest in techniques based on explicit image set matching [9, 16, 25, 26]. This is mainly driven by the need for superior discrimination accuracy as well as increased robustness to practical issues such as pose variations, misalignment and varying environmental conditions (for example, as present in realistic face recognition scenarios [21]).

While image set matching can be accomplished through probability-density based methods [3, 8] and aggregation methods [17], it has been shown that better performance can be attained through modelling image sets via linear structures (ie., subspaces) [25, 29]. Subspaces appear to be appropriate models for this task since they are able to accommodate the effects of various image variations. For example, an acceptable and widely used approximation for photometric invariance, under conditions of no shadowing and Lambertian reflectance, is a 4 dimensional linear space [1].

A convenient way of dealing with subspaces is to represent them as points on Grassmannian manifolds [11, 13, 19, 25]. Recently, several studies explored the applicability of discriminant analysis (DA) on such manifolds [13, 26]. Given subspaces that are represented as points on a Grassmannian manifold $\mathcal{M}$, the underlying idea is to map them to another Grassmannian manifold $\mathcal{M}'$, such that a measure of discriminatory power on $\mathcal{M}'$ is maximised (see Fig. 1 for a conceptual example).

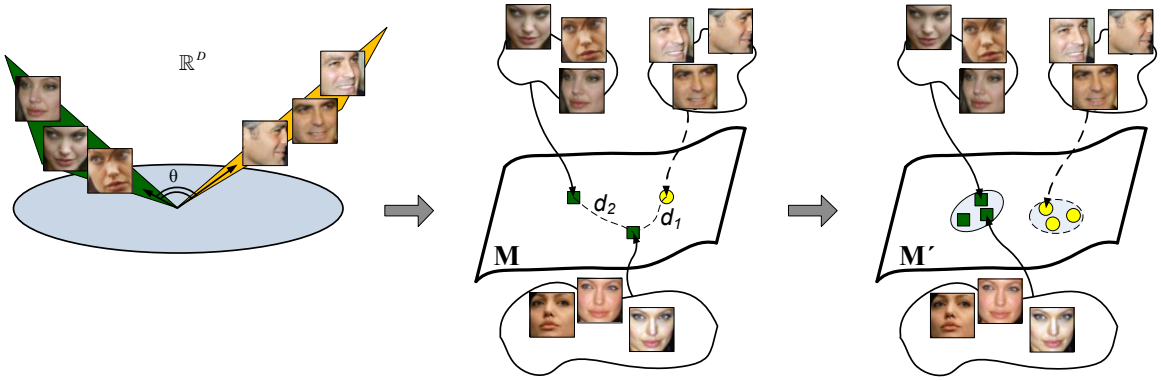
While the approaches presented in [13, 26] show promising results, the conventional formalism of DA suffers from not being able to take into account the local structure of data [10, 15]. For example, outliers and multi-modal classes can adversely affect the discrimination and/or generalisation ability of models based on conventional DA.

Motivated by advances in DA over Euclidean vector spaces [30, 24], we propose a novel DA on Grassmannian manifolds, based on a graph-embedding framework [30]. We show that considerable gains in discrimination accuracy can be obtained by exploiting the geometrical structure and local information on Grassmannian manifolds. This is achieved by introducing within-class and between-class similarity graphs to characterise intra-class compactness and inter-class separability, respectively.

The proposed method for DA on Grassmannian manifolds is somewhat related to distance metric learning methods [28]. The main points of difference include the use of graphs and manifolds in contrast to the typical use of vector spaces in distance metric learning. Overall, the proposed method can be considered as an extension of both graph-embedding and distance metric learning to higher order data structures.

We also propose a new kernel, based on canonical correlation between subspaces, for measuring the similarity of two points on a Grassmannian manifold. We empirically show that, in combination with previous Grassmannian kernels, the new kernel can result in considerable discrimination accuracy improvements.

*Acknowledgements: NICTA is funded by the Australian Government as represented by the *Department of Broadband, Communications and the Digital Economy*, as well as the Australian Research Council through the *ICT Centre of Excellence* program. The second and third authors contributed equally. We thank Prof. Terry Caelli for useful discussions.



(a) **(b)** **(c)**

Figure 1. A conceptual illustration of the proposed approach. **(a)** Image-sets can be described in $\mathbb{R}^D$ by linear subspaces. To compare two linear subspaces, the principal angles between them can be used. For clarity just two subspaces are shown. **(b)** Linear subspaces in $\mathbb{R}^D$ can be represented as points on the Grassmannian manifold $\mathcal{M}$. Having a proper geodesic distance between the points on the manifold, it is possible to convert the image-set matching problem into a point to point classification problem. **(c)** By having a Grassmannian kernel in hand, points on the Grassmannian manifold can be mapped into another Grassmannian manifold where not only certain local properties have been retained but also the discriminatory power between classes has been increased. Unlike the conventional formalism of discriminant analysis, the proposed method preserves the geometrical structure and local information on Grassmannian manifolds by exploiting within-class and between-class similarity graphs.

We continue the paper as follows. Section 2 provides an overview of Grassmannian analysis, which leads to the proposed graph embedding discriminant analysis in Section 3. We introduce the Grassmannian canonical correlation kernel in Section 4. In Section 5 we briefly describe the overall computational complexity of the proposed method. In Section 6 we compare the performance of the proposed method and kernel with previous approaches on several object and face datasets. The main findings and possible future directions are summarised in Section 7.

2. Grassmannian Analysis

Manifold analysis has been extensively considered with success by various disciplines. Amari and Nagaoka state that *many important structures in information theory and statistics can be treated as structures in differential geometry by regarding a space of probabilities as a Riemannian manifold* [2]. A manifold is a topological space that is locally similar to Euclidean space. At an intuitive level, manifolds can be thought of as smooth, curved surfaces embedded in higher dimensional Euclidean spaces. Riemannian manifolds are endowed with a distance measure which allows us to measure how similar two points are. In this work we are interested in a particular class of Riemannian manifolds, known as Grassmannian manifolds [11].

Points on a Grassmannian manifold, $G_{D,m}$, can be viewed as the set of m-dimensional subspaces of $\mathbb{R}^D$ and are represented by orthonormal matrices, each with a size of $D \times m$. Two points on a Grassmannian manifold are equivalent if one can be mapped into the other one by a $m \times m$ orthogonal matrix [11].

Grassmannian analysis provides a natural way to tackle the problem of image set matching. Specifically, as $G_{D,m}$ is the manifold parameterising m-dimensional real vector subspaces of the D-dimensional vector space $\mathbb{R}^D$, the classification problem of matching sets comprising m images, where each image is described by D pixels, can be transformed to a point classification problem on $G_{D,m}$.

During the past decade the concept of angles between subspaces, ie., principal angles has been widely used for image set matching [29]. Since Grassmannian manifolds are curved and the shortest distance between points is geodesic, it is not surprising to see that distances over Grassmannian manifolds may outperform methods based on principal angles. We note that principal angles can be considered as a simple form of geodesic distance on Grassmannian manifolds [19].

Grassmannian kernels [13, 14, 27] allow us to treat the Grassmannian space as if it were a Euclidean vector space. As a result, learning algorithms in vector spaces can be extended to their counterparts on Grassmannian manifolds, eg., kernel discriminant analysis [13, 26]. In the following section we will demonstrate how Grassmannian kernels can be employed to map points on a Grassmannian manifold onto another Grassmannian manifold, where a measure of discriminatory power between classes has been maximised.

3. Graph Embedding Discriminant Analysis

Linear Discriminant Analysis (LDA) is a supervised statistical learning method that seeks a linear projection by simultaneously maximising the between-class dissimilarities and minimising the within-class dissimilarities [6]. While LDA has been successfully applied to various computer vision problems, eg., face recognition [5], it suffers from not being able to naturally capture the local structure of data [10, 24]. For example, LDA has problems handling

multi-modal classes (where each class is comprised of several separate clusters) or when there are outliers in the data. This stems from treating all data points in the same manner (during the calculation of within-class and between-class scatter matrices), no matter how they are related to their classes.

To alleviate the above problem, a graph-embedding framework can be used [7, 24, 30]. A graph $(\boldsymbol{V}, \boldsymbol{W})$ in our context refers to a collection of vertices or nodes, $\boldsymbol{V}$, and a collection of edges that connect pairs of vertices. We note that $\boldsymbol{W}$ is a symmetric matrix with elements describing the similarity between pairs of vertices. Moreover, the diagonal matrix $\boldsymbol{D}$ and the Laplacian matrix $\boldsymbol{L}$ of a graph are defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, with the diagonal elements of $\boldsymbol{D}$ obtained as $\boldsymbol{D}(i,i) = \sum_{j \neq i} \boldsymbol{W}(i,j)$.

Given a graph in a vector space, the purpose of graph-embedding DA is to maximise a measure of discriminatory power by mapping the underlying data into another vector space (usually with lower dimensionality) while preserving similarities between vertex pairs. This problem can be solved through a generalised eigen-analysis framework [30]. In the following text, we formulate the discriminant analysis over Grassmannian manifolds based on the graph-embedding framework.

Given N labelled points $\mathbb{X} = \{(\boldsymbol{X}_i, l_i)\}_{i=1}^N$ from the underlying Grassmannian manifold $\mathcal{M}$, where $\boldsymbol{X}_i \in \mathbb{R}^{D \times m}$ and $l_i \in \{1, 2, \cdots, C\}$, with C denoting the number of classes, the local geometrical structure of $\mathcal{M}$ can be modelled by building a within-class similarity graph $\boldsymbol{W}_w$ and a between-class similarity graph $\boldsymbol{W}_b$. The simplest forms of $\boldsymbol{W}_w$ and $\boldsymbol{W}_b$ are based on the nearest neighbour graphs defined in Eqns. (1) and (2):

$$\boldsymbol{W}_w(i,j) = \begin{cases} 1, & \text{if } \boldsymbol{X}_i \in N_w(\boldsymbol{X}_j) \text{ or } \boldsymbol{X}_j \in N_w(\boldsymbol{X}_i) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$\boldsymbol{W}_b(i,j) = \begin{cases} 1, & \text{if } \boldsymbol{X}_i \in N_b(\boldsymbol{X}_j) \text{ or } \boldsymbol{X}_j \in N_b(\boldsymbol{X}_i) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

In Eqn. (1), $N_w(\boldsymbol{X}_i)$ is the set of v neighbours $\{\boldsymbol{X}_i^1, \boldsymbol{X}_i^2, ..., \boldsymbol{X}_i^v\}$, sharing the same label as l_i. Similarly in Eqn. (2), $N_b(\boldsymbol{X}_i)$ contains v neighbours having different labels. We note that more complex similarity graphs, like heat kernel graphs, can also be used to encode distances between points on Grassmannian manifolds [20].

Our aim is to maximise discriminatory power while simultaneously preserving geometry, by mapping the points on $\mathcal{M}$ to a new manifold $\mathcal{M}'$, ie., $\alpha : \boldsymbol{X}_i \rightarrow \boldsymbol{Y}_i$. A suitable transform would place the connected points of $\boldsymbol{W}_w$ as close as possible, while moving the connected points of $\boldsymbol{W}_b$ as far as possible. Such a mapping can be described by optimising the following two objective functions:

$$f_1 = \min \frac{1}{2} \sum\nolimits_{i,j} (\boldsymbol{Y}_i - \boldsymbol{Y}_j)^2 W_w(i,j) \quad (3)$$

$$f_2 = \max \frac{1}{2} \sum\nolimits_{i,j} (\boldsymbol{Y}_i - \boldsymbol{Y}_j)^2 W_b(i,j) \quad (4)$$

Eqn. (3) punishes neighbours in the same class if they are mapped far away in $\mathcal{M}'$, while Eqn. (4) punishes points of different classes if they are mapped close together in $\mathcal{M}'$. Assume that points on the manifold are implicitly known and only a measure of similarity between them is available through a Grassmannian kernel[1], $k_{ij} = \langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle$.

Confining the solution to be linear, ie., $\boldsymbol{\alpha}_i = \sum_{j=1}^N a_{ij} \boldsymbol{X}_j$, we will have:

$$\boldsymbol{Y}_i = (\langle \boldsymbol{\alpha}_1, \boldsymbol{X}_i \rangle, \langle \boldsymbol{\alpha}_2, \boldsymbol{X}_i \rangle, \cdots, \langle \boldsymbol{\alpha}_r, \boldsymbol{X}_i \rangle)^T \quad (5)$$

By defining $\boldsymbol{A}_l = (a_{l1}, a_{l2}, \cdots, a_{lN})^T$ and $\boldsymbol{K}_i = (k_{i1}, k_{i2}, \cdots, k_{iN})^T$ it can be shown that $\langle \boldsymbol{\alpha}_l, \boldsymbol{X}_i \rangle = \boldsymbol{A}_l^T \boldsymbol{K}_i$. Hence Eqn. (3) can be simplified to:

$$\begin{aligned} & \tfrac{1}{2} \textstyle\sum_{i,j} (\boldsymbol{Y}_i - \boldsymbol{Y}_j)^2 W_w(i,j) \\ = \; & \tfrac{1}{2} \textstyle\sum_{i,j} \left(\boldsymbol{A}_i^T \boldsymbol{K}_i - \boldsymbol{A}_j^T \boldsymbol{K}_j\right)^2 W_w(i,j) \\ = \; & \textstyle\sum_i \boldsymbol{A}_i^T \boldsymbol{K}_i \boldsymbol{K}_i^T \boldsymbol{A}_i^T W_w(i,i) \\ & - \textstyle\sum_{i,j} \boldsymbol{A}_i^T \boldsymbol{K}_j \boldsymbol{K}_i^T \boldsymbol{A}_i^T W_w(i,j) \\ = \; & \mathbb{A}^T \mathbb{K} \boldsymbol{D}_w \mathbb{K}^T \mathbb{A} - \mathbb{A}^T \mathbb{K} \boldsymbol{W}_w \mathbb{K}^T \mathbb{A} \end{aligned} \quad (6)$$

where $\mathbb{A} = [\boldsymbol{A}_1 | \boldsymbol{A}_2 | \cdots | \boldsymbol{A}_r]$ and $\mathbb{K} = [\boldsymbol{K}_1 | \boldsymbol{K}_2 | \cdots | \boldsymbol{K}_N]$. Considering that $\boldsymbol{L}_b = \boldsymbol{D}_b - \boldsymbol{W}_b$, in a similar manner it can be shown that Eqn. (4) can be simplified to:

$$\begin{aligned} & \tfrac{1}{2} \textstyle\sum_{i,j} (\boldsymbol{Y}_i - \boldsymbol{Y}_j)^2 W_b(i,j) \\ = \; & \mathbb{A}^T \mathbb{K} \boldsymbol{D}_b \mathbb{K}^T \mathbb{A} - \mathbb{A}^T \mathbb{K} \boldsymbol{W}_b \mathbb{K}^T \mathbb{A} \\ = \; & \mathbb{A}^T \mathbb{K} \boldsymbol{L}_b \mathbb{K}^T \mathbb{A} \end{aligned} \quad (7)$$

Following [7, 30], a constraint is imposed on Eqn. (3) and the minimisation problem is converted to a maximisation one. Specifically, by forcing $\mathbb{A}^T \mathbb{K} \boldsymbol{D}_w \mathbb{K}^T \mathbb{A}$ to be a constant such as 1, Eqn. (3) becomes the following maximisation problem:

$$\begin{aligned} & \min \left\{ \mathbb{A}^T \mathbb{K} \boldsymbol{D}_w \mathbb{K}^T \mathbb{A} - \mathbb{A}^T \mathbb{K} \boldsymbol{W}_w \mathbb{K}^T \mathbb{A} \right\} \\ = \; & \min \left\{ 1 - \mathbb{A}^T \mathbb{K} \boldsymbol{W}_w \mathbb{K}^T \mathbb{A} \right\} \\ = \; & \max \left\{ \mathbb{A}^T \mathbb{K} \boldsymbol{W}_w \mathbb{K}^T \mathbb{A} \right\} \end{aligned} \quad (8)$$

subject to

$$\mathbb{A}^T \mathbb{K} \boldsymbol{D}_w \mathbb{K}^T \mathbb{A} = 1 \quad (9)$$

By converting both problems into maximisation, the overall optimisation problem is hence:

$$\begin{aligned} & \max \left\{ (\mathbb{A}^T \mathbb{K} (\boldsymbol{L}_b + \beta \boldsymbol{W}_w) \mathbb{K}^T \mathbb{A} \right\} \\ & \text{subject to } \mathbb{A}^T \mathbb{K} \boldsymbol{D}_w \mathbb{K}^T \mathbb{A} = 1 \end{aligned} \quad (10)$$

where β is a Lagrangian multiplier that acts as a regularisation parameter in the final solution. The solution of (10) can be found through the following generalised eigenvalue problem:

$$\mathbb{K} \left\{ \boldsymbol{L}_b + \beta W_w \right\} \mathbb{K}^T \mathbb{A} = \lambda \mathbb{K} D_w \mathbb{K}^T \mathbb{A} \quad (11)$$

More specifically, the desired projection matrix $\mathbb{A}$, is equal to the r largest eigenvectors of the Rayleigh quotient:

$$\frac{\mathbb{K} \boldsymbol{D}_w \mathbb{K}^T}{\mathbb{K} \left\{ \boldsymbol{L}_b + \beta \boldsymbol{W}_w \right\} \mathbb{K}^T} \quad (12)$$

[1] We use the notation $\langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle$ to indicate a similarity measure between points $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ on a Grassmannian manifold. This is similar in principle to an inner product in Hilbert space, as used in kernel-based methods [22].

Fig. 2 outlines the proposed graph embedding method on Grassmannian manifolds. The proposed algorithm uses the points on the Grassmannian manifold implicitly (ie., via measuring similarities through a kernel) to obtain a mapping, $\mathbb{A} = [\boldsymbol{A}_1|\boldsymbol{A}_2|\cdots|\boldsymbol{A}_r]$ that maximises a quotient similar to discriminant analysis, while retaining the overall geometrical structure.

Upon acquiring the mapping $\mathbb{A}$, the matching problem over Grassmannian manifolds is reduced to classification in vector spaces. More precisely, for any query image set $\boldsymbol{X}_q$, a vector representation using the kernel function and the mapping $\mathbb{A}$ is acquired, ie., $\boldsymbol{V}_q = \mathbb{A}^T\boldsymbol{K}_q$, where $\boldsymbol{K}_q = (\langle \boldsymbol{X}_1, \boldsymbol{X}_q\rangle, \langle \boldsymbol{X}_2, \boldsymbol{X}_q\rangle, \cdots, \langle \boldsymbol{X}_N, \boldsymbol{X}_q\rangle)^T$. Similarly, gallery points $\boldsymbol{X}_i$ are represented by r dimensional vectors $\boldsymbol{V}_i = \mathbb{A}^T\boldsymbol{K}_i$ and classification methods such as Nearest-Neighbour or Support Vector Machines [6] can be employed to label $\boldsymbol{X}_q$.

4. Grassmannian Kernels

The similarity between two points on a Grassmannian manifold, eg., $\boldsymbol{X}_i$ and $\boldsymbol{X}_j \in \mathbb{R}^{D\times m}$, can be measured using kernels such as the projection kernel:

$$k_{i,j}^{[\mathrm{proj}]} = \left\|\boldsymbol{X}_i^T\boldsymbol{X}_j\right\|_F^2 \tag{13}$$

One of the first attempts to solve the problem of image set matching was based on the notion of *principal angles*. More precisely, Yamaguchi et al. [29] used the largest canonical correlation value (the cosine of principal angles) to measure the similarity between two image sets. In Section 4.1 we show that the largest canonical correlation between subspaces is a kernel on Grassmannian manifolds. We then show in Section 4.2 that a more complex kernel, created through linearly combining existing Grassmannian kernels, is also a Grassmannian kernel.

We will later demonstrate that combining the projection kernel with the proposed canonical correlation kernel can lead to considerable improvements in discrimination accuracy, in the context of the proposed graph-embedding discriminant analysis.

4.1. Canonical Correlation Kernel

Given subspaces $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, we define the canonical correlation kernel as:

$$k_{i,j}^{[\mathrm{CC}]} = \max_{\boldsymbol{a}_p\in\mathrm{span}(\boldsymbol{X}_i)}\ \max_{\boldsymbol{b}_q\in\mathrm{span}(\boldsymbol{X}_j)} \boldsymbol{a}_p^T\boldsymbol{b}_q \tag{14}$$

subject to $\boldsymbol{a}_p^T\boldsymbol{a}_p = \boldsymbol{b}_p^T\boldsymbol{b}_p = 1$ and $\boldsymbol{a}_p^T\boldsymbol{a}_q = \boldsymbol{b}_p^T\boldsymbol{b}_q = 0,\ p \neq q$.

For $k^{[\mathrm{CC}]}$ to be a Grassmannian kernel [14], it must be **(i)** positive definite, and **(ii)** well defined, meaning it is invariant to various representations of the subspaces, ie., $k(\boldsymbol{X}_1, \boldsymbol{X}_2) = k(\boldsymbol{X}_1\boldsymbol{R}_1, \boldsymbol{X}_2\boldsymbol{R}_2),\ \forall\ \boldsymbol{R}_1, \boldsymbol{R}_2 \in Q(m)$, where $Q(m)$ indicates orthonormal matrices of order m.

Since the singular values of $\boldsymbol{X}_1^T\boldsymbol{X}_2$ are equal to $\boldsymbol{R}_1^T\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{R}_2$, the canonical correlation kernel is well-defined. To show that the kernel matrix $[\mathbb{K}]_{ij} = k_{i,j}^{[\mathrm{CC}]}$ is positive definite, it suffices to show that $z^T\mathbb{K}z > 0$ for $\forall z \in \mathbb{R}^n$:

$$\begin{aligned} z^T\mathbb{K}z &= \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}^T \begin{pmatrix} k_{1,1}^{[\mathrm{CC}]} & k_{1,2}^{[\mathrm{CC}]} & \dots & k_{1,n}^{[\mathrm{CC}]} \\ k_{2,1}^{[\mathrm{CC}]} & k_{2,2}^{[\mathrm{CC}]} & \dots & k_{2,n}^{[\mathrm{CC}]} \\ \vdots & \vdots & & \vdots \\ k_{n,1}^{[\mathrm{CC}]} & k_{n,2}^{[\mathrm{CC}]} & \dots & k_{n,n}^{[\mathrm{CC}]} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \\ &= z_1^2k_{1,1}^{[\mathrm{CC}]} + z_2^2k_{2,2}^{[\mathrm{CC}]} + \ldots + z_n^2k_{n,n}^{[\mathrm{CC}]} \\ &\quad +2\left(z_1z_2k_{1,2}^{[\mathrm{CC}]} + z_1z_3k_{1,3}^{[\mathrm{CC}]} + \ldots + z_1z_nk_{1,n}^{[\mathrm{CC}]}\right) \\ &\quad +2\left(z_2z_3k_{2,3}^{[\mathrm{CC}]} + z_2z_4k_{2,4}^{[\mathrm{CC}]} + \ldots + z_2z_nk_{2,n}^{[\mathrm{CC}]}\right) \\ &\quad + \ldots + 2z_{n-1}z_nk_{n-1,n}^{[\mathrm{CC}]} \end{aligned} \tag{15}$$

In Eqn. (15) we have used the fact that $k_{i,j}^{[\mathrm{CC}]} = k_{j,i}^{[\mathrm{CC}]}$. Since the principal angle between $\boldsymbol{X}_i$ to itself is zero, $k_{i,i}^{[\mathrm{CC}]} = 1$. Hence Eqn. (15) can be further simplified to:

$$\begin{aligned} z^T\mathbb{K}z &= \left(\sum_{i=1}^n z_i\right)^2 - 2\sum_{i=1}^n\sum_{j\neq i} z_iz_j + 2\sum_{i=1}^n\sum_{j\neq i} z_iz_jk_{i,j}^{[\mathrm{CC}]} \\ &= \left(\sum_{i=1}^n z_i\right)^2 + 2\sum_{i=1}^n\sum_{j\neq i} z_iz_j\left(k_{i,j}^{[\mathrm{CC}]} - 1\right) \end{aligned} \tag{16}$$

Note that $\min\left(z_iz_j\left(k_{i,j}^{[\mathrm{CC}]}\text{-}1\right)\right) = -z_iz_j$, since $k_{i,j}^{[\mathrm{CC}]} \in [0,1]$. Consequently:

$$\min\left(z^T\mathbb{K}z\right) = \left(\sum_{i=1}^n z_i\right)^2 - 2\sum_{i=1}^n\sum_{j\neq i} z_iz_j \tag{17}$$

As the right-hand side of Eqn. (17) is always positive for $z_i \neq 0$, $\mathbb{K}$ is a positive-definite matrix.

Input:

- Training set $\mathbb{X} = \{(\boldsymbol{X}_i, l_i)\}_{i=1}^N$ from the underlying Grassmannian manifold, where $\boldsymbol{X}_i \in \mathbb{R}^{D\times m}$ is a subspace (obtained for example via SVD over an image-set) and $l_i \in \{1, 2, \cdots, C\}$, with C denoting the number of classes
- A kernel function k_{ij}, for measuring the similarity between two points on the Grassmannian manifold

Processing:

1. Compute the Gram matrix $[\mathbb{K}]_{ij}$ for all $\boldsymbol{X}_i, \boldsymbol{X}_j$
2. Compute the within-class and between-class graph similarity matrices, $\boldsymbol{W}_w$, $\boldsymbol{W}_b$, respectively, the between Laplacian matrix $\boldsymbol{L}_b$ and the diagonal within matrix $\boldsymbol{D}_w$
3. To obtain $\mathbb{A}$, solve the maximisation problem in Eqn. (11) by eigen decomposition; $\mathbb{A}$ is equal to the r largest eigenvectors of the Rayleigh quotient $\frac{\mathbb{K}\boldsymbol{D_w}\mathbb{K}^T}{\mathbb{K}\{\boldsymbol{L_b}+\beta\boldsymbol{W_w}\}\mathbb{K}^T}$

Output:

- The projection matrix $\mathbb{A} = [\boldsymbol{A}_1|\boldsymbol{A}_2|\cdots|\boldsymbol{A}_r]$, where each $\boldsymbol{A}_i$ is an eigenvector found in step 3 above; the eigenvectors are sorted in a descending manner according to their corresponding eigenvalues

---

Figure 2. Pseudocode for training Grassmannian graph-embedding discriminant analysis.

4.2. Grassmannian Kernel Combinations

In general, we can express a linear combination of two Grassmannian kernels $k^{[\mathtt{A}]}$ and $k^{[\mathtt{B}]}$ as:

$$k^{[\mathtt{A}+\mathtt{B}]} = \gamma^{[\mathtt{A}]} k^{[\mathtt{A}]} + \gamma^{[\mathtt{B}]} k^{[\mathtt{B}]} \tag{18}$$

where $\gamma^{[\mathtt{A}]}, \gamma^{[\mathtt{B}]} \geq 0$. From the theory of Reproducing Kernel Hilbert Space (RKHS) we know that the superposition of two kernels is a new kernel [22]. As such and in order to extent the superposition rule over Grassmannian manifold, it suffices to show that the superposition kernel is well-defined. Since,

$$\begin{aligned} & k^{[\mathtt{A}]}\left(\boldsymbol{R}_1\boldsymbol{X}_1, \boldsymbol{R}_2\boldsymbol{X}_2\right) && + k^{[\mathtt{B}]}\left(\boldsymbol{R}_1\boldsymbol{X}_1, \boldsymbol{R}_2\boldsymbol{X}_2\right) \\ = \quad & k^{[\mathtt{A}]}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) && + k^{[\mathtt{B}]}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) \end{aligned} \tag{19}$$

$k^{[\mathtt{A}]} + k^{[\mathtt{B}]}$ is well-defined and Eqn. (18) depicts a valid Grassmannian kernel.

5. Computational Complexity

The solution to (12) is found using Singular Value Decomposition (SVD), which has the computational complexity of $O(s^3)$ for a square matrix of size $s \times s$. Solving the generalised eigenvector problem hence demands for $O(N^3)$ operations. Computing $k^{[\mathtt{CC}]}$ and $k^{[\mathtt{proj}]}$ demands for $O(\frac{N(N-1)}{2}(m^2D+m^3))$ and $O(\frac{N(N-1)}{2}m^2D))$, respectively. Considering that $m << D$ and $N << D$, the computational complexity of the proposed algorithm is hence $O(m^2DN^2)$.

6. Experiments

The proposed approach[2] was compared and contrasted to previous state-of-the-art methods on two image set recognition tasks: face and object recognition. We will first briefly overview the datasets used in the experiments (Section 6.1), followed by a description and discussion of the experiments (Section 6.2).

6.1. Setup of Image Datasets

For the face recognition task we used three datasets: CMU-PIE [23], BANCA [4] and CMU-MoBo [12]. For the object recognition task we used the ETH-80 dataset [18]. For all datasets we randomly split the images into training and test sets. 10 random splits were obtained. Following [9, 13, 27, 29], we used normalised pixel intensities as image features and generated subspaces using SVD.

CMU-PIE contains images of 68 people captured under 13 poses, 43 illuminations conditions, and with 4 expressions. In our experiments, near frontal poses (c05, c07, c09, c27, c29) were used. See Fig. 3 for examples of the variations. We generated 180 image sets as training data and 300 image sets as test data. Images were cropped to the internal part of the face (ie., closely cropped, no background) and downsampled to 32×32 pixels.

Figure 3. Examples of appearance variations in **(a)** CMU-PIE, **(b)** BANCA, **(c)** CMU-MoBo. The variations include pose, illumination, expression and image quality.

BANCA contains image sets for 52 people (26 male and 26 female). For each person video recordings were made under various conditions (illumination, pose and camera variations), while the person was talking. In each condition two recordings were made per person and 5 images were extracted from each video. We generated 150 image sets as training data and 150 sets as test data. All faces were closely cropped and resized to 64×64.

CMU-MoBo consists of motion sequences of 25 people walking on a treadmill. For each person video recordings were made for 4 walking styles (slow walk, fast walk, inclined walk and slow walk while holding a ball), viewed from a set of fixed cameras. We generated 72 image sets as training data and 216 sets as test data. Images were cropped and normalised to 32×32.

ETH-80 contains images of eight object categories: apples, cows, cups, dogs, horses, pears, tomatoes, and cars. Each category includes ten object subcategories (eg., various dogs) in 41 orientations, resulting in 410 images per category. Examples are shown in Fig. 4. We resized the images to 64×64. Unlike the face images mentioned above, the background was kept. We generated 24 image sets as gallery data and 56 sets as probe data.

6.2. Performance Comparison

The proposed algorithm was compared against: standard geodesic distance on Grassmannian manifolds, Grassmann Discriminant Analysis (GDA) [13], Kernel Grassmannian Discriminant Analysis (KGDA) [26], and the kernel version

[2] Matlab/Octave source code for the proposed method is available at *http://itee.uq.edu.au/~uqmhara1*

Figure 4. **(a)** examples from the eight object categories in the ETH-80 dataset; **(b)** examples of various classes within an object category.

of Affine Hull Image-Set Distance (Kernel AHISD) [9]. For GDA and KGDA the projection kernel ($k^{[\mathtt{proj}]}$) was used. For the first kernel in KGDA (the kernel applied in vector space) we considered Gaussian, polynomial and linear kernels; the best obtained results are reported.

We acknowledge that the graph parameter v and size of projection matrix r must be duly adjusted in the proposed approach. The graph parameter v is dependent on the number of samples per class and distribution of points over Grassmannian manifold. Our empirical studies suggest that $v \in [1, 10]$ provides satisfactory results; however, the optimal value can be determined by searching over a range of possible values. In the following experiments we have used the maximum number of eigenvectors $r = N - 1$ for deriving the projection matrix.

The average recognition accuracy and standard deviation across the 10 random splits of each dataset are reported. For each split of the dataset, two evaluations were done, each using a different number of images per set. For example, we used sets with 6 and 9 images for the BANCA dataset. For the sake of simplicity, the classification scheme in all experiments was nearest-neighbour [6].

Table 1 shows the results for geodesic distance, GDA, KGDA, as well as the proposed algorithm in conjunction with $k^{[\mathtt{proj}]}$. The results indicate that the proposed algorithm obtains the highest recognition accuracy, in all bar one case (CMU-MoBo with 6 images per set). We note that the proposed algorithm outperforms geodesic distance and GDA by a significant margin for all the tests[3]. Moreover, the proposed method also considerably outperforms KGDA on all datasets except CMU-MoBo.

The superior performance of geodesic distance over GDA and KGDA is an implication of a relatively difficult recognition task, and can be intuitively explained by the global behaviour of DA. Specifically, DA might underperform if the underlying data cannot be modelled effectively by a Gaussian distribution.

[3]We also considered the less demanding ETH-80 experiment setup used in [13] and obtained 100% accuracy with the projection kernel.

Table 2 shows the results for Kernel AHISD and the proposed algorithm in combination with three kernels: **(i)** projection kernel $k^{[\mathtt{proj}]}$, **(ii)** canonical correlation kernel $k^{[\mathtt{CC}]}$, **(iii)** combined kernel $k^{[\mathtt{proj+CC}]}$.

For $k^{[\mathtt{proj+CC}]}$, based on Eqn. (18), the mixing coefficient $\gamma^{[\mathtt{proj}]}$ was fixed at 1, while the optimal value of $\gamma^{[\mathtt{CC}]}$ was found by scanning through a range of values. The results do not seem to vary much as long as $\gamma^{[\mathtt{CC}]}$ is large enough (approximately 5 to 10). The range of values for $k^{[\mathtt{proj}]}$ was found to be typically around 0.5 to 3, while $k^{[\mathtt{CC}]}$ has a maximum value of 1 (see Section 4.1). As such, $\gamma^{[\mathtt{CC}]}$ acts as a scaling factor, making the contribution of $\gamma^{[\mathtt{proj}]}$ and $\gamma^{[\mathtt{CC}]}$ to $k^{[\mathtt{proj+CC}]}$ roughly comparable. Fig. 5 shows the performance for various values of $\gamma^{[\mathtt{CC}]}$.

The proposed algorithm, in conjunction with $k^{[\mathtt{proj+CC}]}$, considerably outperforms Kernel AHISD on all datasets except CMU-MoBo, where the two approaches obtain quite similar results. The results for the proposed algorithm also show that neither $k^{[\mathtt{proj}]}$ or $k^{[\mathtt{CC}]}$ dominates. In some cases $k^{[\mathtt{proj}]}$ is better than $k^{[\mathtt{CC}]}$, while in others the reverse is true, suggesting that the two kernels are more suited to different data distributions. By combining the two kernels, ie., $k^{[\mathtt{proj+CC}]}$, noticeably better results than either of the two kernels are obtained. This further suggests that $k^{[\mathtt{proj}]}$ and $k^{[\mathtt{CC}]}$ are describing different aspects, which in turn suggests that the proposed canonical correlation kernel can be useful.

The results in Tables 1 and 2 also indicate that using more images helps in most cases. However, for the proposed method in conjunction with $k^{[\mathtt{CC}]}$, the opposite appears to occur. Though it cannot be stated conclusively that either kernel is more suitable for larger sets[4], a possible explanation could be the violation of the linearity assumption as used in modelling of subspaces. Similar behaviour can be observed for Kernel AHISD, where a similar modelling assumption is used.

[4]To explore this further, we performed an extra experiment on CMU-MoBo using 24 images per set. In this case $k^{[\mathtt{CC}]}$ outperformed $k^{[\mathtt{proj}]}$ by more than 10 percentage points.

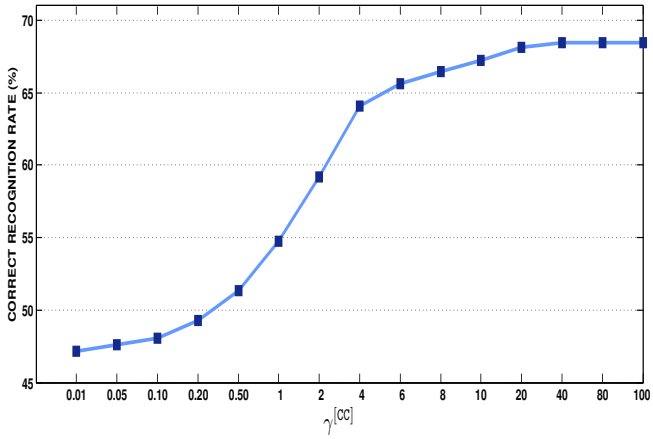


Figure 5. Performance on CMU-PIE for various values of $\gamma^{[\mathtt{CC}]}$, used for combining $k^{[\mathtt{proj}]}$ and $k^{[\mathtt{CC}]}$.

Table 1. Average correct recognition rate for image set matching using geodesic distance, GDA [13], KGDA [26], as well as the proposed algorithm in conjunction with the projection kernel. The standard deviation is shown in brackets. The suffix in the dataset name indicates the number of images per set (eg. the 6 in BANCA-6).

Dataset	Geodesic	GDA [13]	KGDA [26]	Proposed algorithm with $k^{[\mathtt{proj}]}$, eqn. (13)
CMU-PIE-6	33.70 (4.9)	24.07 (9.8)	20.23 (4.8)	**42.67** (5.4)
CMU-PIE-15	57.80 (3.5)	53.33 (4.4)	50.00 (3.7)	**65.27** (4.4)
BANCA-6	48.00 (4.4)	46.60 (2.9)	39.80 (4.5)	**53.27** (4.5)
BANCA-9	60.13 (1.8)	56.80 (3.2)	50.07 (3.6)	**64.53** (2.3)
CMU-MoBo-6	55.52 (2.9)	58.53 (2.8)	**64.68** (2.3)	61.24 (2.5)
CMU-MoBo-9	61.00 (1.6)	63.65 (2.5)	64.36 (1.4)	**64.90** (1.7)
ETH-80-6	85.71 (3.7)	83.21 (5.7)	67.50 (5.9)	**90.89** (2.8)
ETH-80-15	85.17 (3.4)	84.11 (4.9)	63.21 (5.7)	**91.25** (2.1)

Table 2. Average correct recognition rate for image set matching using Kernel AHISD [9], as well as the proposed algorithm in conjunction with various kernels. The standard deviation is shown in brackets.

Dataset	Kernel AHISD [9]	Proposed algorithm with $k^{[\mathtt{proj}]}$, eqn. (13)	Proposed algorithm with $k^{[\mathtt{CC}]}$, eqn. (14)	Proposed algorithm with $k^{[\mathtt{proj+CC}]}$, eqn. (18)
CMU-PIE-6	46.60 (6.6)	42.67 (5.4)	52.93 (5.6)	**68.47** (6.8)
CMU-PIE-15	44.66 (4.2)	65.27 (4.4)	53.67 (1.3)	**75.80** (1.6)
BANCA-6	19.87 (1.2)	53.27 (4.5)	56.33 (1.9)	**63.00** (2.5)
BANCA-9	19.33 (0.6)	64.53 (2.3)	55.00 (1.4)	**68.73** (1.6)
CMU-MoBo-6	**87.72** (1.5)	61.24 (2.5)	78.81 (2.9)	86.27 (1.3)
CMU-MoBo-9	89.70 (1.1)	64.90 (1.7)	76.62 (2.5)	**89.92** (1.8)
ETH-80-6	69.11 (5.1)	90.89 (2.8)	75.71 (7.6)	**91.96** (3.1)
ETH-80-15	68.21 (7.9)	91.25 (2.1)	73.93 (4.5)	**92.32** (2.4)

7. Main Findings and Future Directions

In this paper we have proposed a novel image set matching approach, based on Grassmannian manifolds. Specifically, our approach employs a graph-embedding framework and derives a mapping on Grassmannian manifolds to simultaneously maximise a measure of discriminatory power and preserve the geometrical structure of the manifold. We have also introduced a new Grassmannian kernel, based on canonical correlation between subspaces.

When compared to several recent methods (GDA, KGDA, Kernel AHISD), experiments on several image datasets suggest that the proposed approach can obtain considerable improvements in discrimination accuracy. The experiments also show that the new kernel, when used in combination with the projection kernel, leads to further increases in accuracy.

When the new and projection kernels were separately evaluated for comparison purposes, there was no clear cut winner — on some datasets the projection kernel was bet-

ter, while on others the proposed kernel was better. As such, a more comprehensive study may provide more insights as to why the two kernels are more suited to particular datasets.

Future avenues of research include exploring subset generation prior to Grassmannian analysis. More precisely, by clustering a set of images into several subsets and considering each subset as a point on a Grassmannian manifold, richer descriptions on Grassmannian manifolds might be attained.

References

[1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *IEEE Tran. Pattern Analysis and Machine Intelligence*, 19(7):721 –732, 1997.

[2] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society, 2001.

[3] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. In *Proc. Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 581–588, 2005.

[4] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. The BANCA database and evaluation protocol. In *Audio- and Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Computer Science (LNCS)*, volume 2688, pages 625–638, 2003.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997.

[6] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] D. Cai, X. He, K. Zhou, J. Han, and H. Bao. Locality sensitive discriminant analysis. In *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, pages 708–713, 2007.

[8] F. Cardinaux, C. Sanderson, and S. Bengio. User authentication via adapted statistical models of face images. *IEEE Trans. Signal Processing*, 54(1):361–373, 2006.

[9] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 2567–2573, 2010.

[10] J. Chen, J. Ye, and Q. Li. Integrating global and local structures: A least squares framework for dimensionality reduction. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.

[11] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[12] R. Gross and J. Shi. The CMU Motion of Body (MoBo) Database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.

[13] J. Hamm and D. D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proc. Int. Conf. Machine Learning (ICML)*, pages 376–383, 2008.

[14] J. Hamm and D. D. Lee. Extended Grassmann kernels for subspace-based learning. In *Neural Information Processing Systems (NIPS)*, pages 601–608, 2009.

[15] M. Harandi, M. Nili Ahmadabadi, and B. Araabi. Optimal local basis: A reinforcement learning approach for face recognition. *International Journal of Computer Vision*, 81(2):191–204, 2009.

[16] T.-K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(6):1005–1018, 2007.

[17] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(3):226–239, 1998.

[18] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 409–415, 2003.

[19] Y. M. Lui, J. Beveridge, B. Draper, and M. Kirby. Image-set matching using a geodesic distance and cohort normalization. In *Proc. Automatic Face Gesture Recognition (AFGR)*, pages 1–6, 2008.

[20] S. Rosenberg. *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds*. Cambridge University Press, 1997.

[21] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. *Lecture Notes in Computer Science (LNCS)*, Vol. 5558, pages 199–208, 2009.

[22] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[23] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12):1615–1618, 2003.

[24] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.*, 8:1027–1061, 2007.

[25] R. Wang, S. Shan, X. Chen, and W. Gao. Manifold-manifold distance with application to face recognition based on image set. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[26] T. Wang and P. Shi. Kernel grassmannian distances and discriminant analysis for face recognition from image sets. *Pattern Recognition Letters*, 30(13):1161–1165, 2009.

[27] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *J. Mach. Learn. Res.*, 4:913–931, 2003.

[28] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. J. Russell. Distance metric learning with application to clustering with side-information. In *Neural Information Processing Systems (NIPS)*, pages 505–512, 2002.

[29] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proc. Automatic Face and Gesture Recognition (AFGR)*, pages 318–323, 1998.

[30] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(1):40–51, 2007.