

Kernel Sparse Subspace Clustering on Symmetric Positive Definite Manifolds

Ming Yin¹, Yi Guo², Junbin Gao³, Zhaoshui He¹ and Shengli Xie¹

¹School of Automation, Guangdong University of Technology, Guangzhou 510006, P. R. China

²School of Computing, Engineering and Mathematics, Western Sydney University, Parramatta, NSW 2150, Australia

³The University of Sydney Business School, The University of Sydney, Camperdown, NSW 2006, Australia

{yiming, zhshhe, shlxie}@gdut.edu.cn; y.guo@westernsydney.edu.au; junbin.gao@sydney.edu.au

Abstract

Sparse subspace clustering (SSC), as one of the most successful subspace clustering methods, has achieved notable clustering accuracy in computer vision tasks. However, SSC applies only to vector data in Euclidean space. Unfortunately there is still no satisfactory approach to solve subspace clustering by self-expressive principle for symmetric positive definite (SPD) matrices which is very useful in computer vision. In this paper, by embedding the SPD matrices into a Reproducing Kernel Hilbert Space (RKHS), a kernel subspace clustering method is constructed on the SPD manifold through an appropriate Log-Euclidean kernel, termed as kernel sparse subspace clustering on the SPD Riemannian manifold (KSSCR). By exploiting the intrinsic Riemannian geometry within data, KSSCR can effectively characterize the geodesic distance between SPD matrices to uncover the underlying subspace structure. Experimental results on several famous datasets demonstrate that the proposed method achieves better clustering results than the state-of-the-art approaches.

1. Introduction

Despite the fact that the majority of subspace clustering methods [23] show good performance in various applications, the similarity among data points is measured in the original data domain. Specifically, this similarity is often measured between the sparse [5] or low-rank representations [14] of data points, by exploiting the so-called *self-expressive* property of the data [6], in terms of Euclidean distance or alike. In general, the representation for each data point is a vector of its linear regression coefficients on the rest of the data subject to sparsity or low-rank constraint. Unfortunately, this assumption may not be always true for many high-dimensional data in real world where data may be better modeled by nonlinear manifolds [9, 11, 16]. In this case, the *self-expressive* based algorithms, such as Sparse Subspace Clustering (SSC) [6] and Low Rank Sub-

space Clustering (LRSC) [24], are no longer applicable. For example, the human facial images can be regarded as samples from a nonlinear submanifold [12, 17].

Recently, a useful image and video descriptor, the covariance descriptor which is a symmetric positive definite (SPD) matrix [21], has attracted a lot of attention. By using this descriptor, a promising classification performance can be achieved [9]. However, the traditional subspace learning mainly focuses on the problem associated with vectorial data. It is known that SPD matrices form a Lie group, a well structured Riemannian manifold. The naive way of vectorizing SPD matrices first and applying any of the available vector-based techniques usually seems less intuitive and short of proper interpretation. The underlying reason is the lack of vector space structures in Riemannian manifold. That is, the direct application of linear reconstruction model for this type of data will result in inaccurate representation and hence compromised performance.

A few solutions have been recently proposed to address sparse coding problems on Riemannian manifolds, such as [3, 10, 19]. While for subspace clustering, a nonlinear LRR model is proposed to extend the traditional LRR from Euclidean space to Stiefel manifold [29], SPD manifold [7] and abstract Grassmann manifold [26] respectively. The common strategy employed in these work is to use the logarithm mapping “projecting” data onto the tangent space at each data point where a “normal” Euclidean linear reconstruction of a given sample is well defined [22] so that the self-expressive principle works. This idea was first explored by Ho *et al.* [10] and they proposed a nonlinear generalization of sparse coding to handle the non-linearity of Riemannian manifolds, by flattening the manifold using a fixed tangent space.

For Riemannian manifold of SPD, many researchers [3, 19] took advantage of a nice property of this manifold, namely its closeness under positive linear combination, and exploited appropriate nonlinear metrics such as Log-determinant divergence to measure errors in the sparse model formulation. This type of strategies fall in the second

category of methods dealing with problems on manifolds by embedding manifold onto a large flat Euclidean space. For example, Shirazi *et. al*[30] embedded the Grassmann manifolds into Reproducing Kernel Hilbert Spaces to improve the clustering of data (such as image sequences). Furthermore, the LRR model on abstract Grassmann manifold is recently proposed based on the embedding technique [27] and its kernelization [25].

It is noteworthy that a nonlinear extension of SSC method for manifold clustering has been proposed in [17]. Unfortunately, the authors simply used the kernel trick to map data onto a high-dimensional feature space without considering the intrinsic geometric structure. To rectify this, in this paper, we propose to use a kernel feature mapping to embed the SPD Riemannian manifold into a high dimension feature space and preserve its intrinsic Riemannian geometry. We call this method kernel sparse subspace clustering on Riemannian manifold (KSSCR). An overview of the proposed method is illustrated in Fig. 1. Different from the work in [17], our motivation is to map SPD matrices into Reproducing Kernel Hilbert Space (RKHS) with Log-Euclidean Gaussian kernel based on Riemannian metric. As a result, the linear reconstruction can be naturally implemented in the feature space associated with the kernel where the original SSC can be applied. The proposed method not only effectively characterizes the geodesic distance between pair of SPD matrices but also uncovers the underlying low-dimensional subspace structure.

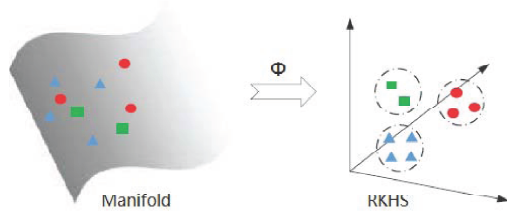


Figure 1: Overview of the proposed kernel sparse subspace clustering method. Using the SPD kernel, data is mapped onto a high-dimensional RKHS so as to attain better representations for clustering.

The remainder of the paper is organized as follows. In Section 2, we give a brief review on related work. Section 3 is dedicated to introducing the novel kernel sparse subspace clustering on Riemannian manifold. The experimental results are given in Section 4 and Section 5 concludes the paper with a brief summary.

2. Related Work

Before introducing the proposed model, in this section, we briefly review the recent development of subspace clustering methods [6] and the analysis of Riemannian geom-

etry of SPD Manifold [18]. Throughout the paper, capital letters denote matrices (e.g., X) and bold lower-case letters denote column vectors (e.g., \mathbf{x}). x_i is the i -th element of vector \mathbf{x} . Similarly, X_{ij} denotes the (i, j) -th entry of matrix X . $\|\mathbf{x}\|_1 = \sum_i |x_i|$ and $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$ are the ℓ_1 and ℓ_2 norms respectively, where T is the transpose operation. $\|\cdot\|_F$ is the matrix Frobenius norm defined as $\|X\|_F = \sqrt{\sum_{ij} |X_{ij}|^2}$. The manifold of $d \times d$ SPD matrices is denoted by \mathcal{S}_d^+ . The tangent space at a point X on \mathcal{S}_d^+ is defined by $T_X \mathcal{S}_d^+$, which is a vector space including the tangent vectors of all possible curves passing through X .

2.1. Subspace Sparse Representation

Sparse representation, which has proved to be a useful tool for representing and compressing high-dimensional signals, provides a statistical model for finding efficient and compact signal representations. Among them, SSC [6] is of particular interests as it clusters data points to a union of low-dimensional subspaces without referring to a library. Let $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ be a matrix of data. Each datum \mathbf{x}_i is drawn from a low-dimensional subspace denoted by \mathcal{S}_j for $j \in [1, \dots, k]$. By exploiting the *self-expressive* property of the data, the formulation of SSC is written as follows,

$$\min_{\mathbf{C}} \|\mathbf{C}\|_1 + \lambda \|\mathbf{E}\|_F^2 \text{ s.t. } X = X\mathbf{C} + \mathbf{E}, \text{diag}(\mathbf{C}) = 0. \quad (1)$$

where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N] \in \mathbb{R}^{N \times N}$ is the coefficient matrix whose column \mathbf{c}_i ¹ is the sparse representation vector corresponding to the i -th data point. \mathbf{E} denotes the reconstruction error components.

As to the coefficient matrix \mathbf{C} , besides the interpretation as new sparse representation of the data, each element C_{ij} in \mathbf{C} can also be regarded as a similarity measure between the data pair \mathbf{x}_i and \mathbf{x}_j . In this sense, \mathbf{C} is sometimes called an affinity matrix. Therefore, a clustering algorithm such as K-means can be subsequently applied to \mathbf{C} for the final segmentation solution. This is a common practice of subspace clustering based on finding new representation.

2.2. Sparse Representation on SPD matrices

Since SPD matrices belong to a Lie group which is a Riemannian manifold [1], it cripples many methods that rely on linear reconstruction. Generally, there are two methods to deal with the non-linearity of Riemannian manifolds. One is to locally flatten the manifold to tangent spaces[22]. The underlying idea is to exploit the geometry of the manifold directly. The other is to map the data into a feature space usually a Hilbert space [11]. Precisely, it is to project the data into RKHS through kernel mapping [8]. Both of these

¹Matrix \mathbf{C} is bold, while other matrices are not.

methods are seeking a transformation so that the linearity emerges.

A typical example of the former method is the one in [10]. Let X be a SPD matrix and hence a point on \mathcal{S}_d^+ . $\mathbb{D} = \{D_1, D_2, \dots, D_N\}$, $D_i \in \mathcal{S}_d^+$ is a dictionary. An optimization problem for sparse coding of X on a manifold \mathcal{M} is formulated as follows

$$\min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \left\| \sum_{i=1}^N w_i \log_X(D_i) \right\|_X^2, \quad \text{s.t.} \quad \sum_{i=1}^N w_i = 1, \quad (2)$$

where $\log_X(\cdot)$ denotes Log map from SPD manifold to a tangent space at X , $\mathbf{w} = [w_1, w_2, \dots, w_N]$ is the sparse vector and $\|\cdot\|_X$ is the norm associated with $T_X \mathcal{S}_d^+$. Because $\log_X(X) = \mathbf{0}$, the second term in Eq.(2) is essentially the error of linearly reconstructing $\log_X(X)$ by others on the tangent space of X . As this tangent space is a vector space, this reconstruction is well defined. As a result, the traditional sparse representation model can be performed on Riemannian manifold.

However, it turns out that quantifying the reconstruction error is not at all straightforward. Although ℓ_2 -norm is commonly used in the Euclidean space, using Riemannian metrics would be better in \mathcal{S}_d^+ since they can accurately measure the intrinsic distance between SPD matrices. In fact, a natural way to measure closeness of data on a Riemannian manifold is geodesics, i.e. curves analogous to straight lines in \mathbb{R}^n . For any two data points on a manifold, geodesic distance is the length of the shortest curve on the manifold connecting them. For this reason, the affine invariant Riemannian metric (AIRM) is probably the most popular Riemannian metric defined as follows [18]. Given $X \in \mathcal{S}_d^+$, the AIRM of two tangent vectors $\mathbf{v}, \mathbf{w} \in T_X \mathcal{S}_d^+$ is defined as

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \langle X^{-1/2} \mathbf{v} X^{-1/2}, X^{-1/2} \mathbf{w} X^{-1/2} \rangle \\ &= \text{tr}(X^{-1} \mathbf{v} X^{-1} \mathbf{w}). \end{aligned}$$

The geodesic distance between points $X, Y \in \mathcal{S}_d^+$ induced from AIRM is then

$$\delta_g(X, Y) = \|\log(X^{-1/2} Y X^{-1/2})\|_F. \quad (3)$$

where $\log(\cdot)$ is the principal matrix logarithm operator.

3. Kernel Subspace Clustering on SPD Matrices

Motivated by solving the above problems, in this section, we propose a novel kernel sparse subspace clustering algorithm which enables SSC to handle data on Riemannian manifold by incorporating the intrinsic geometry of the manifold. The idea is quite simple but effective: map the data points into RKHS first and then perform SSC with some

modifications in RKHS. Compared with the original SSC, the advantages of this approach include simpler solutions and better representation due to the capability of learning the underlying nonlinear structures. The following is the detail of our method. Given a data set $\mathcal{X} = [X_1, X_2, \dots, X_N]$ on SPD manifold, we seek its sparse representation via exploiting the *self-expressive* property of the data. Thus, the objective of our kernel sparse subspace representation algorithm on Riemannian manifold is formulated as follows

$$\begin{aligned} \min_C \lambda \|C\|_1 + \sum_{i=1}^N \left\| \phi(X_i) - \sum_{j=1}^N c_{ij} \phi(X_j) \right\|_F^2, \\ \text{s.t.} \quad \text{diag}(C) = 0. \end{aligned} \quad (4)$$

where $\phi(\cdot)$ denotes a feature mapping function that projects SPD matrices into RKHS such that $\langle \phi(X), \phi(Y) \rangle = \kappa(X, Y)$ where $\kappa(X, Y)$ is a positive definite (PD) kernel.

3.1. Log-Euclidean Kernels for SPD Matrices

Although locally flattening Riemannian manifolds via tangent spaces [10] can handle their non-linearity, it inevitably leads to very demanding computation due to switching back and forth between tangent spaces and the manifold. Furthermore, linear reconstruction of SPD matrices is not as natural as in Euclidean space and this may incur errors.

The recent work in [9] shows that the Stein divergence is akin to AIRM. Furthermore, a PD kernel can be induced from Stein divergence under some conditions [20]. Concretely, a Stein metric [20], also known as Jensen-Bregman LogDet divergence (JBLD) [4], derived from Bregman matrix divergence is given by,

$$\delta_s(X, Y) = \log \left| \frac{X+Y}{2} \right| - \frac{1}{2} \log |XY|,$$

where $|\cdot|$ denotes determinant. Accordingly a kernel function based on Stein divergence for SPD \mathcal{S}_d^+ can be defined as $\kappa_s(X, Y) = \exp\{-\beta \delta_s(X, Y)\}$, though it is guaranteed to be positive definite only when $\beta \in \{\frac{1}{2}, 1, \dots, \frac{d-1}{2}\}$ or $\beta > \frac{d-1}{2}$ [20].

However, the problem is that Stein divergence is only an approximation to Riemannian metric and cannot be a PD kernel without more restricted conditions [12]. If one uses this kernel, the reconstruction error resulting from representing X_i with other X_j under Riemannian metric will incur. To address this problem, a family of Log-Euclidean kernels were proposed in [12], i.e., a polynomial kernel, an exponential kernel and a Gaussian kernel, tailored to model data geometry more accurately. These Log-Euclidean kernels were proven to well characterize the true geodesic distance between SPD matrices, especially the Log-Euclidean Gaussian kernel

$$\kappa_g(X, Y) = \exp\{-\gamma \|\log(X) - \log(Y)\|_F^2\},$$

which is a PD kernel for any $\gamma > 0$. Owing to its superiority, we select Log-Euclidean Gaussian kernel to transform the SPD matrices into RKHS.

3.2. Optimization

In this subsection, we solve the objective of kernel sparse subspace learning in (4) via alternating direction method of multipliers (ADMM) [2]. Expanding the Frobenius norm in (4) and applying the kernel trick [8] leads to the following equivalent problem

$$\begin{aligned} \min_{\mathbf{C}} \quad & \lambda \|\mathbf{C}\|_1 - 2\text{tr}(\mathbf{K}\mathbf{C}) + \text{tr}(\mathbf{C}\mathbf{K}\mathbf{C}^T), \\ \text{s.t.} \quad & \text{diag}(\mathbf{C}) = 0, \end{aligned} \quad (5)$$

where $\mathbf{K} = \{k_{ij}\}_{i,j=1}^n$ is the kernel Gram matrix, and $k_{ij} = \kappa_g(X_j, X_i) = \phi(X_j)^T \phi(X_i)$.

By introducing an auxiliary matrix \mathbf{A} , the above problem can be rewritten as follows

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{A}} \quad & \lambda \|\mathbf{C}\|_1 - 2\text{tr}(\mathbf{K}\mathbf{A}) + \text{tr}(\mathbf{A}\mathbf{K}\mathbf{A}^T), \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{C} - \text{diag}(\mathbf{C}). \end{aligned} \quad (6)$$

The augmented Lagrangian is then

$$\begin{aligned} \mathcal{L}(\mathbf{C}, \mathbf{A}) = \min_{\mathbf{C}, \mathbf{A}} \quad & \lambda \|\mathbf{C}\|_1 - 2\text{tr}(\mathbf{K}\mathbf{A}) + \text{tr}(\mathbf{A}\mathbf{K}\mathbf{A}^T) \\ & + \frac{\rho}{2} \|\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C})\|_F^2 + \text{tr}(\Delta^T (\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C}))), \end{aligned} \quad (7)$$

on which ADMM is carried out. Note that Δ is a Lagrangian multiplier matrix with compatible dimension.

The optimization algorithm is detailed as follows.

1. Update \mathbf{A} .

$$\begin{aligned} \min_{\mathbf{A}} \quad & -2\text{tr}(\mathbf{K}\mathbf{A}) + \frac{\rho}{2} \|\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C})\|_F^2 \\ & + \text{tr}(\mathbf{A}\mathbf{K}\mathbf{A}^T) + \text{tr}(\Delta^T (\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C}))) \end{aligned} \quad (8)$$

Let $\tilde{\mathbf{C}} = \mathbf{C} - \text{diag}(\mathbf{C})$, the subproblem can be formulated by,

$$\begin{aligned} \min_{\mathbf{A}} \quad & -2\text{tr}(\mathbf{K}\mathbf{A}) + \text{tr}(\mathbf{A}\mathbf{K}\mathbf{A}^T) \\ & + \frac{\rho}{2} \|\mathbf{A} - \tilde{\mathbf{C}}\|_F^2 + \text{tr}(\Delta^T (\mathbf{A} - \tilde{\mathbf{C}})) \end{aligned} \quad (9)$$

Setting derivative w.r.t. \mathbf{A} to zero results in a closed-form solution to subproblem (8) given by

$$\mathbf{A}_{t+1} = (2\mathbf{K} + \rho\tilde{\mathbf{C}}_t - \Delta_t)(2\mathbf{K} + \rho\mathbf{I})^{-1}. \quad (10)$$

where \mathbf{I} is an identity matrix.

2. Update \mathbf{C} .

$$\begin{aligned} \min_{\mathbf{C}} \quad & \lambda \|\mathbf{C}\|_1 + \frac{\rho}{2} \|\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C})\|_F^2 \\ & + \text{tr}(\Delta^T (\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C}))). \end{aligned} \quad (11)$$

The above subproblem has the following closed-form solution given by shrinkage operator

$$\mathbf{C}_{t+1} = \mathbf{J} - \text{diag}(\mathbf{J}), \quad (12)$$

$$\mathbf{J} = \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{A}_{t+1} + \frac{\Delta_t}{\rho}). \quad (13)$$

where $\mathcal{S}_{\eta}(\cdot)$ is a shrinkage operator acting on each element of the given matrix, and is defined as $\mathcal{S}_{\eta}(v) = \text{sgn}(v)\max(|v| - \eta, 0)$.

3. Update Δ .

$$\Delta_{t+1} = \Delta_t + \rho(\mathbf{A}_{t+1} - \mathbf{C}_{t+1} + \text{diag}(\mathbf{C}_{t+1})). \quad (14)$$

These steps are repeated until $\|\mathbf{A}_t - \mathbf{C}_t\|_{\infty} \leq \epsilon$, $\|\mathbf{A}_{t+1} - \mathbf{A}_t\|_{\infty} \leq \epsilon$.

3.3. Subspace Clustering

As discussed earlier, \mathbf{C} is actually a new representation learned from data. After solving problem (4), the next step is to segment \mathbf{C} to find the final subspace clusters. Here we apply a spectral clustering method to the affinity matrix given by $(|\mathbf{C}| + |\mathbf{C}^T|)/2$ to separate the data into clusters, which is equivalent to subspaces. We call this algorithm Kernel Subspace Clustering on SPD (KSSCR for short) and it is outlined in Algorithm 1.

Algorithm 1: KSSCR

Input: $\mathcal{X} = [X_1, X_2, \dots, X_N]$, γ , λ , ρ and ϵ .

Steps:

1. Solve (4) by ADMM explained in Section 3.2, and obtain the optimal solution \mathbf{C}^* .
2. Compute the affinity matrix \mathcal{W} by,

$$\mathcal{W} = (|\mathbf{C}^*| + |\mathbf{C}^{*T}|)/2.$$

3. Apply normalized spectral clustering method to \mathcal{W} to obtain the final clustering solution \mathcal{C} .

Output: the clustering solution \mathcal{C} .

3.4. Complexity Analysis and Convergence

The computational cost of the proposed algorithm is mainly determined by the steps in ADMM. The total complexity of KSSCR is, as a function of the number of data

points, $\mathcal{O}(\frac{1}{3}N^3 + tN^2)$ where t is the total number of iterations. The soft thresholding to update the sparse matrix C in each step is relatively cheap, as the complexity is much lower than $\mathcal{O}(N^2)$. For updating A we can pre-compute the Cholesky decomposition of $(2K + \rho\mathbf{I})^{-1}$ at the cost of less than $\mathcal{O}(\frac{1}{2}N^3)$, then compute new A by using (10) which has a complexity of $\mathcal{O}(N^2)$ in general.

The above proposed ADMM iterative procedure to the augmented Lagrangian problem (7) satisfies the general condition for the convergence theorem in [13].

4. Experimental Results

In this section, we present several experimental results to demonstrate the effectiveness of KSSCR. To comprehensively evaluate the performance of KSSCR, we tested it on texture images, human faces and person appearance. Some sample images of test databases are shown in Figure 2. The clustering results are shown in Section 4.1, Section 4.2 and 4.3 respectively. All test data lie on Riemannian manifold.

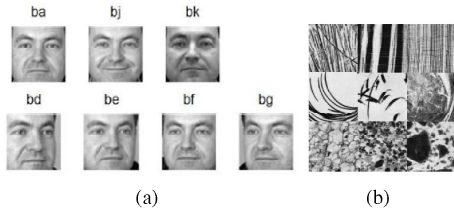


Figure 2: Samples on the FERET (a) and Brodatz (b) database.

In this work, we adopt two criteria, i.e. Normalized Mutual Information (NMI) and subspace clustering accuracy, to quantify the clustering performance more precisely. Meanwhile, to extensively assess how the proposed algorithm improves the performance of data clustering, the following four state-of-the-art subspace clustering methods are compared against:

1. Sparse Subspace Clustering(SSC) [6].
2. Low-rank Representation(LRR) [15].
3. Low-Rank Subspace Clustering (LRSC) [24], which seeks a low-rank representation by decomposing the corrupted data matrix as the sum of a clean and self-expressive dictionary.
4. Kernel SSC on Euclidean space (KSSCE) [17], which embeds data onto to a nonlinear manifold by using the kernel trick and then apply SSC based on Euclidean metric.

We also included K-means clustering algorithm (K-means) as a baseline.

4.1. Texture Clustering

In this subsection, a subset of the Brodatz database [12], i.e., 16-texture ('16c') mosaic, was chosen for clustering performance evaluation. There are 16 objects in this subset in which each class contains only one image. Before clustering, we downsampled each image to 256×256 and then split into 64 regions of size 32×32 . To obtain their region covariance matrices (RCM), a feature vector $f(x, y)$ for any pixel $I(x, y)$ was extracted, e.g., $f(x, y) = (I(x, y), |\frac{\partial I}{\partial x}|, |\frac{\partial I}{\partial y}|, |\frac{\partial^2 I}{\partial x^2}|, |\frac{\partial^2 I}{\partial y^2}|)$. Then, each region can be characterized by a 5×5 covariance descriptor. Totally, 1024 RCM were collected. We randomly chose some data from Brodatz database with the number of clusters, i.e., N_c , ranging from 2 to 16. The final performance scores were computed by averaging the scores from 20 trials. The detailed clustering results are summarized in Tables 1 and 2. We set the parameters as $\lambda = 0.04$ and $\gamma = 0.5$ for KSSCR. Also, the tuned parameters are reported for the results achieved by other methods. The bold numbers highlight the best results.

From the tables, we observe that the proposed method outperforms other methods in most cases while KSSCE achieved the second best performance. This is due to the nonlinear subspace clustering in Euclidean space without using Riemannian metric. In addition, we find SSC type methods can better discover the intrinsic structure than LRR type ones. In order to show the underlying low-dimensional structure within data, we provide a visual comparison of affinity matrices obtained by different methods in Figure 3. Due to the fact that LRSC cannot recover the subspace effectively, we exclude its affinity matrix. It is clear that the affinity matrix obtained by KSSCR effectively reflects the structure of data so as to benefit the subsequent clustering task.

4.2. Face Clustering

In this test, we used the "b" subset of FERET database which consists of 1400 images with the size of 80×80 from 200 subjects (7 images each subject). The images of each individuals were taken under different expression and illumination conditions, marked by 'ba', 'bd', 'be', 'bf', 'bg', 'bj', and 'bk'. To represent a facial image, similar to [28], we created a 43×43 region covariance matrix, i.e., a specific SPD matrix, which is summarised over intensity value, spatial coordinates, 40 Gabor filters at 8 orientations and 5 scales.

We first tested KSSCR on seven subsets from FERET database which randomly covers some different clusters. The clustering results of the tested approaches are shown in Figure 4 and 5 with varying number of clusters. From Figure 4, we observe that the clustering accuracy of the proposed approach is better than other state-of-the-art methods

N_c	2	4	8	10	12	16
K-means	73.44±0	63.28±0	76.95±0	61.88±0	59.24±0	60.45±0
SSC (100)	100±0	79.30±0	78.03±5.18	61.77±0.86	68.27±3.26	60.27±1.15
LRSC(1, 0.01)	53.90±0	33.98±0	22.63±0.30	21.88±0.50	16.87±0.33	15.95±0.46
LRR(0.4)	71.88±0	42.58±0	52.17±1.66	53.52±0.22	50.77±0.81	48.34±1.84
KSSCE(0.4)	78.91 ± 0	53.52 ± 0	77.64 ± 0.93	72.36 ± 0.77	82.81±0.65	64.93 ± 2.89
KSSCR	100±0	85.94±0	97.85±0	74.64±0.08	75.7±0.04	83.62±0.14

Table 1: Clustering results in terms of accuracy (%) on Brodatz database.

N_c	2	4	8	10	12	16
K-means	28.70 ± 0	45.49 ± 0	75.76 ± 0	63.76 ± 0	62.27 ± 0	63.88 ± 0
SSC(100)	100±0	63.84±0	80.21±2.60	64.17 ±0.30	70.65±1.93	67.03±0.58
LRSC(1, 0.01)	0.44 ± 0	2.08±0	10.86±0.4	14.17 ±0.26	10.23 ± 0.43	15.37±0.43
LRR(0.4)	21.24±0	15.15 ± 0	50.58 ± 1.65	48.32 ±0.36	54.21 ± 0.62	55.14 ± 0.89
KSSCE(0.4)	28.65 ± 0	35.95 ± 0	73.52 ± 0.53	70.35 ± 0.63	79.42±0.08	66.35 ± 2.06
KSSCR	100±0	73.62±0	95.26±0	76.71±0.07	79.05±0.04	82.38±0.11

Table 2: Clustering results in terms of NMI (%) on Brodatz database.

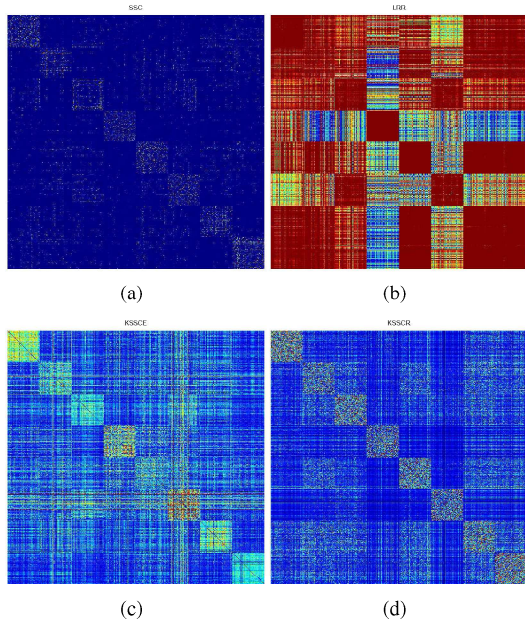


Figure 3: Examples of affinity matrices on 8-cluster data from the Brodatz database (a) SSC, (b) LRR, (c) KSSCE and (d) KSSCR.

in most cases, peaking on 30-cluster subset with $\lambda = 0.1$ and $\gamma = 2 \times 10^{-3}$. On average, the clustering rate on all seven subsets are 44.49, 49.33, 78.56, 47.54, 72.50, and 80.42 for K-means, SSC, LRSC, LRR, KSSCE, and KSSCR, respectively. In Figure 5, the NMI is presented to show the

performance of different methods. The average scores are 66.65, 68.31, 88.43, 69.08, 83.01, and 88.46 for K-means, SSC, LRSC, LRR, KSSCE, and KSSCR, respectively. As can be seen, KSSCR achieves the comparable results and a little bit better than that of LRSC. While compared to KSSCE, KSSCR is leading by a large margin. This further verifies the advantage of fully considering the intrinsic geometry structure within data.

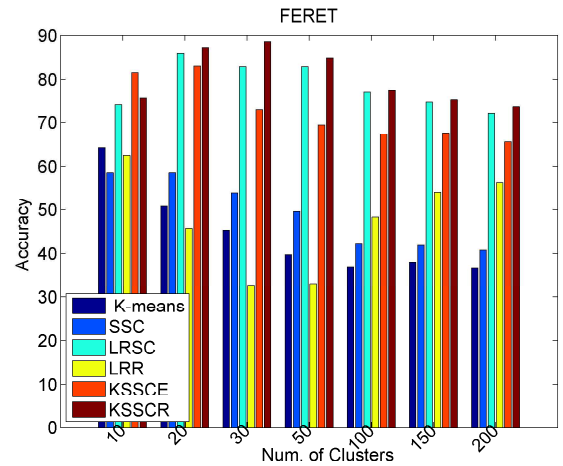


Figure 4: Comparison of clustering accuracy (%) on FERET dataset. The average scores on all 7 tests are 44.49, 49.33, 78.56, 47.54, 72.50, and 80.42 for K-means, SSC, LRSC, LRR, KSSCE, and KSSCR, respectively.

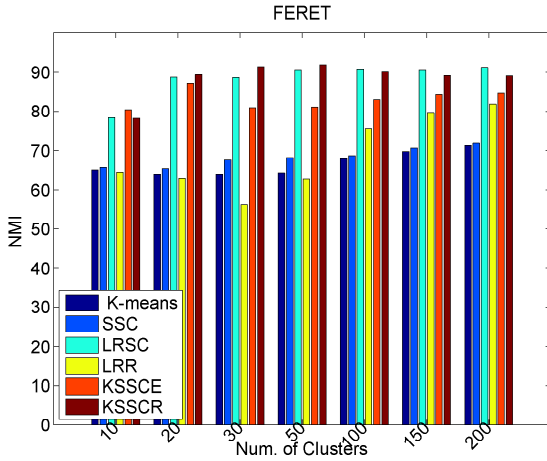


Figure 5: Comparison of NMI (%) on FERET dataset. The average scores on all 7 tests are 66.65, 68.31, 88.43, 69.08, 83.01, and 88.46 for K-means, SSC, LRSC, LRR, KSSCE, and KSSCR, respectively.

Next, we tested the effect of the hyper-parameter γ in Log-Euclidean Gaussian kernel by fixing the number of clusters to 30. Fig. 6 shows the clustering performance versus parameter γ on the FERET dataset. We can see that the clustering score increases as γ gets larger, reaching peak at about 2×10^{-3} and decreasing afterwards. This helps to determine the value of γ in clustering experiments.

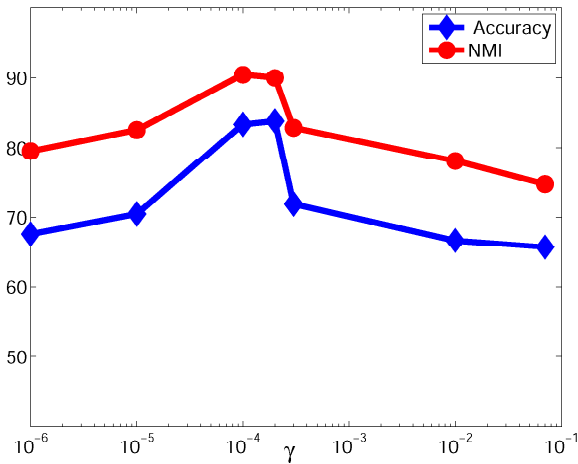


Figure 6: Accuracy and NMI (%) (y-axis) of KSSCR with different γ (x-axis) on FERET dataset.

4.3. Appearance clustering

Finally, we present the person appearance clustering results obtained by various methods. Here, we used the Se-

Methods	ACC	NMI
K-means	45.68 ± 0	68.62 ± 0
SSC (100)	52.79 ± 1.04	74.57 ± 0.52
LRSC (5,100)	51.01 ± 1.10	73.12 ± 0.53
LRR (0.369)	54.17 ± 1.03	75.42 ± 0.62
KSSCE (0.05)	53.60 ± 1.62	74.80 ± 0.52
KSSCR	65.77 ± 0.01	81.06 ± 0.01

Table 3: Clustering results in terms of accuracy and NMI (%) on the ETHZ pedestrian Seq.1 database including 80 classes.

quence 1 of the modified ETHZ dataset[31] which is the most challenging one. It contains 80 pedestrians (4,857 images). In order to facilitate the subsequent processing, we prepared the covariance descriptors as in [9] so that each region can be depicted by a 17×17 covariance matrix.

We selected the parameters $\lambda = 0.03$ and $\gamma = 0.05$ for our approach. For the sake of fairness, we tuned the parameters for other compared methods to achieve their best results as well. The clustering results are reported in Table3. As can be seen, the proposed KSSCR achieves the best score against other state-of-the-art methods.

5. Conclusion

In this paper, we proposed a novel algorithm called kernel sparse subspace clustering (KSSCR) for SPD matrices, a special type of data that originally cripples original sparse representation methods based on linear reconstruction. By using an appropriate Log-Euclidean kernel, we exploited the *self-expressive* property to obtain sparse representation of the original data mapped into kernel reproducing Hilbert space based on Riemannian metric. Experimental results show that the proposed method can provide better clustering solutions than the state-of-the-art approaches thanks to incorporating Riemannian geometry structure.

Acknowledgement

The Project was supported in part by the Guangdong Natural Science Foundation under Grant (No.2014A030313511), in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, China, in part by Science and Technology Planning Project of Guangdong Province(No.2014B040401012), in part by Science and Technology Planning Project of Guangzhou Province(No.2014J4100204) and by Foundation for University Construction(No.12ZK0362).

References

- [1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328-347, 2007.
- [2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, volume 3. Foundations and Trends in Machine Learning, 2011.
- [3] A. Cherian and S. Sra. Riemannian sparse coding for positive definite matrices. In *Proceedings of ECCV*, volume 8691, pages 299-314. Springer International Publishing, 2014.
- [4] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2161-2174, Sept 2013.
- [5] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6-18, 2006.
- [6] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765-2781, 2013.
- [7] Y. Fu, J. Gao, X. Hong, and D. Tien. Low rank representation on Riemannian manifold of symmetric positive definite matrices. In *Proceedings of SDM*, 2015, DOI:10.1137/1.9781611974010.36.
- [8] M. Harandi, C. Sanderson, R. Hartley, and B. Lovell. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *Proceedings of ECCV*, pages 216-229, 2012.
- [9] M. T. Harandi, R. Hartley, B. C. Lovell, and C. Sanderson. Sparse coding on symmetric positive definite manifolds using bregman divergences. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, DOI:10.1109/TNNLS.2014.2387383.
- [10] J. Ho, Y. Xie, and B. C. Vemuri. On a nonlinear generalization of sparse coding and dictionary learning. In *Proceedings of ICML*, volume 28, pages 1480-1488, 2013.
- [11] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. T. Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In *Proceedings of CVPR*, pages 73-80, June 2013.
- [12] P. Li, Q. Wang, W. Zuo, and L. Zhang. Log-Euclidean kernels for sparse representation and dictionary learning. In *Proceedings of ICCV*, pages 1601-1608, Dec 2013.
- [13] Z. Lin, R. Liu, and H. Li. Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. *Machine Learning*, 99(2):287-325, 2015.
- [14] G. Liu, Z. Lin, S. Yan, J. Sun, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171 - 184, Jan. 2013.
- [15] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of ICML*, pages 663-670, 2010.
- [16] H. Nguyen, W. Yang, F. Shen, and C. Sun. Kernel low-rank representation for face recognition. *Neurocomputing*, 155:32-42, 2015.
- [17] V. M. Patel and R. Vidal. Kernel sparse subspace clustering. In *Proceedings of ICIP*, pages 2849-2853, Oct 2014.
- [18] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal Of Computer Vision*, 66:41-66, 2006.
- [19] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos. Tensor sparse coding for positive definite matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):592-605, 2014.
- [20] S. Sra. A new metric on the manifold of kernel matrices with application to matrix geometric means. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Proceedings of NIPS*, pages 144-152, 2012.
- [21] O. Tuzel, F. Porikli, and P. Meer. Region Covariance: A Fast Descriptor for Detection And Classification. In *Proceedings of ECCV*, pages 589-600, 2006.
- [22] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713-1727, Oct 2008.
- [23] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52-68, 2011.
- [24] R. Vidal and P. Favaro. Low rank subspace clustering (LRSC). *Pattern Recognition Letters*, 43(1):47-61, 2014.
- [25] B. Wang, Y. Hu, J. Gao, Y. Sun, and B. Yin. Kernelized low rank representation on Grassmann manifolds. *arXiv preprint arXiv:1504.01806*.
- [26] B. Wang, Y. Hu, J. Gao, Y. Sun, and B. Yin. Low rank representation on Grassmann manifolds: An extrinsic perspective. *arXiv preprint arXiv:1504.01807*.
- [27] B. Wang, Y. Hu, J. Gao, Y. Sun, and B. Yin. Low rank representation on Grassmann manifold. In *Proceedings of ACCV*, pages 81-96, Singapore, November 2014.
- [28] M. Yang, L. Zhang, S. C. Shiu, and D. Zhang. Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary. *Pattern Recognition*, 46(7):1865 - 1878, 2013.
- [29] M. Yin, J. Gao, and Y. Guo. Nonlinear low-rank representation on Stiefel manifolds. *Electronics Letters*, 51(10):749-751, 2015.
- [30] S. Shirazi, M. T. Harandi, C. Sanderson, A. Alavi, and B. C. Lovell. Clustering on Grassmann manifolds via kernel embedding with application to action analysis. In *Proceedings of ICIP*, 2012, 781-784.
- [31] W. Schwartz, L. Davis. Learning Discriminative Appearance-Based Models Using Partial Least Squares. In *Proceedings of SIBGRAPI*, 2009, 322-329.