Low Rank Representation on Riemannian Manifold of Symmetric Positive Definite Matrices

Yifan Fu [1] Junbin Gao[1] Xia Hong[2] David Tien[1]

[1] School of Computing and Mathematics
Charles Sturt University
Bathurst, NSW 2795, Australia
Email:{yfu, jbgao,dtien}@csu.edu.au

[2] School of Systems Engineering
University of Reading
Reading, RG6 6AY, UK
Email:x.hong@reading.ac.uk

May 1, 2015

Introduction–LRR

- Low rank representation (LRR) is an effective method to explore the intrinsic low rank structures embedded in a data set in high dimensional space, which has been successfully used in many applications, such as motion segmentation , image segmentation and salient object detection.
- Given a collection of data points $\mathcal{X} = \{X_1, \cdots, X_n\}$, LRR seeks a joint low rank representation of $\mathcal{X}$ using data points themselves as the dictionary, which can be formulated as,

$$\min_{W} \|\mathcal{X} - \mathcal{X}W\|^2 + \lambda\|W\|_\star \tag{1.1}$$

where $W = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_n \end{pmatrix}$ is the low rank representation matrix under the dictionary $\mathcal{X}$.

Introduction–Covariance matrices

- However, for many applications, such as those in machine learning , computer vision and medical image analysis, data are not characterized by simple vector features.
  - Taking image data as an example, the raw pixel features , such as color, gradient and filter responses are not robust in the presence of illumination changes and non-rigid motion. To mitigate the variances in raw pixel features, a natural way is to gather statistical information of the raw pixel features.
- The covariance of a set of raw features inside a region of interest is one of the successful feature descriptors, which is called '*covariance matrix*'.
- Covariance matrices as feature descriptors offer a convenient platform for fusing multiple features into a compact form independent of the number of data points.
- Covariance matrices are symmetric positive definite(SPD). It is well-known that all the SPD matrices form the so-called curved Riemannian manifold.

Related work

- Two issues associated with the SPD matrices
  - For the given dictionary atom $\{X_k\}$, there is no guarantee for the linear combination $\sum_{k=1}^{n} w_{ik} X_k$ to be a SPD matrix.
  - The Euclidean metric does not make sense to measure the error between X_k and $\sum_{k=1}^{n} w_{ik} X_k$.
- Current Research on the SPD matrices
  - For the first issue, it can be easily resolved by assuming combination coefficients w_{ik} be non-negative.
  - For the second issue, different non-Euclidean geometry strategies have been proposed for sparse coding, such as the log-determinant divergence, Affine Invariant Riemannian Measures (AIRM) and Kernel functions.
- To our best knowledge, none of existing work is specialized for the low rank representation for SPD matrices measured simultaneously by the Riemmanian distance and LRR, which motivates our study.

Preliminaries

- A *manifold* $\mathcal{M}$ of dimensional d is a topological space that locally resembles a Euclidean space $\mathbb{R}^d$ in a neighbourhood of each point $X \in \mathcal{M}$. For example, lines and circles are $1D$ manifolds, and surfaces, such as a plane, a sphere, and a torus, are $2D$ manifolds.
- A *tangent vector* is a vector that is tangent to a manifold at a given point X. All the possible tangent vectors at X constitutes a Euclidean space, named the *tangent space* of $\mathcal{M}$ at X and denoted by $T_X\mathcal{M}$.
- If we have a smoothly defined metric (inner-product) across all the tangent spaces $\langle \cdot, \cdot \rangle_X : T_X\mathcal{M} \times T_X\mathcal{M} \to \mathbb{R}$ on every point $X \in \mathcal{M}$, then we call $\mathcal{M}$ *Riemannian manifold.*
- There are predominantly two operations for computations on the Riemannian manifold, namely (1) the exponential map at point X, denoted by $\exp_X : T_X\mathcal{M} \to \mathcal{M}$, and (2) the logarithmic map, at point X, $\log_X : \mathcal{M} \to T_X\mathcal{M}$. The former projects a tangent vector in the tangent space onto the manifold, the latter does the reverse.
- The distance between two points $X_i, X_j \in \mathcal{M}$ can be calculated through the following formula as the norm in tangent space:

$$\text{dist}_{\mathcal{M}}(X_i, X_j) = \| \log_{X_i}(X_j) \|_{X_i} \tag{3.1}$$

LRR on Riemannian manifold

- We propose a novel LRR model on the manifold of SPD matrices. The approximation quality is measured by the extrinsic Euclidean distance defined by the metric on tangent spaces.
- The LRR model in Eq. (1.1) can be changed to the following manifold form:

$$\min_{W} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} w_{ij} \log_{X_i}(X_j) \right\|_{X_i}^{2} + \|W\|_{\star} \quad \text{s.t.} \sum_{j=1}^{n} w_{ij} = 1, i = 1, 2, \ldots, n \tag{4.1}$$

The Difference between Euclidean LRR and our new method

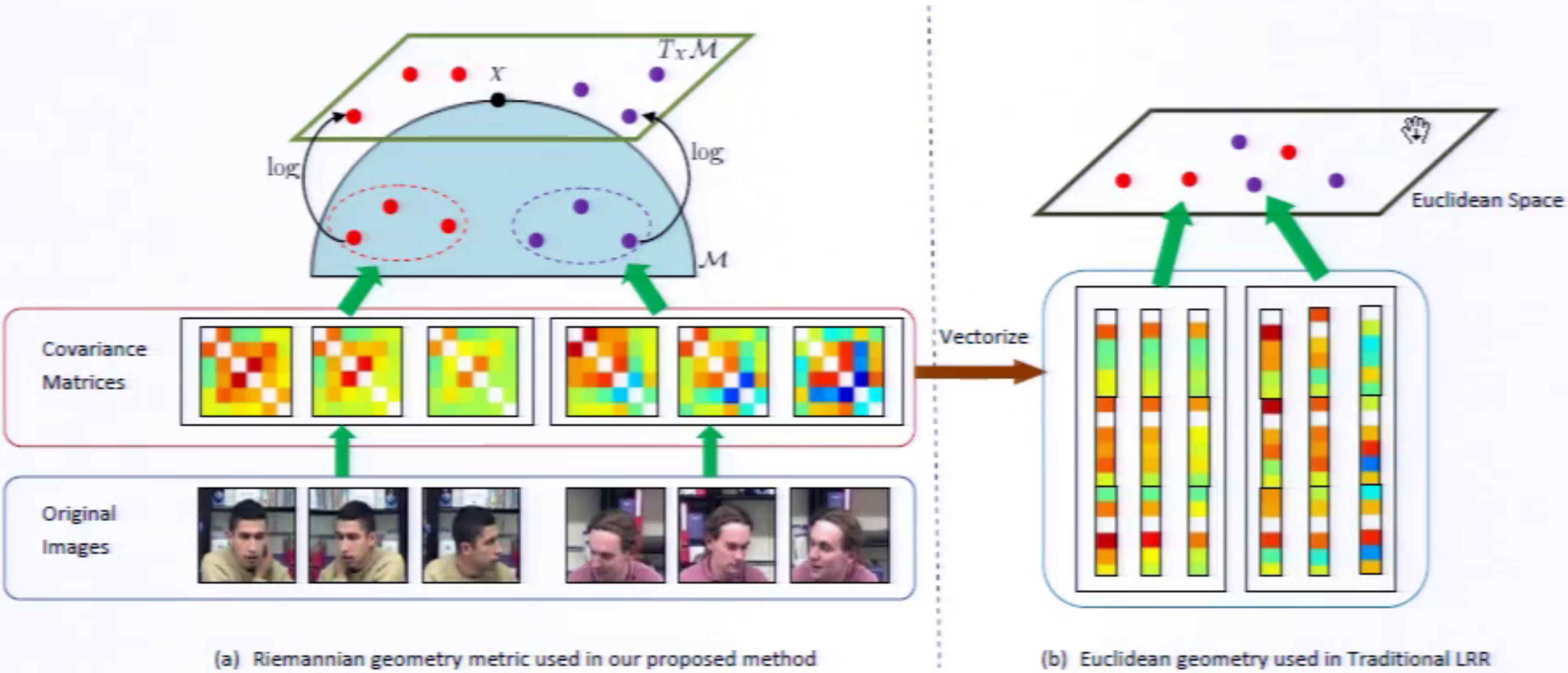


Figure: The illustration of distance metrics used in our proposed method and Euclidean LRR methods

- Given two points $X, Z \in S_+(d)$, their distance on tangent space is formulated by

$$\|\log_X(Z)\|_X^2 = \mathrm{tr}(\mathrm{Log}^2(G^{-1}ZG^{-T})) \tag{4.2}$$

where G denotes the square root matrix of X, i.e., $G = X^{\frac{1}{2}}$.

- Let $L_{ij} = \mathrm{Log}(G_i^{-1}X_jG_i^{-T})$, then Eq.(4.1) can be written into a matrix form as follows:

$$\begin{aligned} &\min_W \frac{1}{2}\sum_{i=1}^n \mathbf{w}_i Q_i \mathbf{w}_i^T + \lambda\|W\|_\star \\ &\text{s.t. } \sum_{j=1}^n w_{ij} = 1, i = 1, 2, \ldots, n \end{aligned} \tag{4.3}$$

where $\mathbf{w}_i$ is the i-th row of W, and $Q_i = [\mathrm{tr}(L_{ij}L_{ik})]$ are $n \times n$ matrices.

Solution to LRR on SPD matrices

- We propose to use the Augmented Lagrange Multiplier (ALM) method to solve the constrained optimization problem in Eq. (4.3).
- First of all, the augmented Lagrange problem of (4.3) can be written as

$$L = \sum_{i=1}^{n} \left(\frac{1}{2} \mathbf{w}_i Q_i \mathbf{w}_i^T + y_i (\sum_{j=1}^{n} w_{ij} - 1) + \frac{\beta}{2} (\sum_{j=1}^{n} w_{ij} - 1)^2 \right) + \lambda ||W||_\star \tag{5.1}$$

  where y_i are Lagrangian multipliers, and β is a weight to tune the error term of $(\sum_{j=1}^{n} w_{ij} - 1)^2$.
- The above problem can be solved by updating one variable at a time with all the other variables fixed. More specifically, the iterations of ALM go as follows:
  - Fix all others to update W.
  - Fix all others to update y_i by

$$y_i^{k+1} \leftarrow y_i^k + \beta_k (\sum_{j=1}^{n} w_{ij}^{k+1} - 1) \tag{5.2}$$

Fix all others to update W

$$L = \sum_{i=1}^{n} \left(\frac{1}{2}\mathbf{w}_i Q_i \mathbf{w}_i^T + y_i(\sum_{j=1}^{n} w_{ij} - 1) + \frac{\beta}{2}(\sum_{j=1}^{n} w_{ij} - 1)^2 \right) + \lambda ||W||_\star \tag{5.3}$$

$$L = F(W) + \lambda ||W||_\star \tag{5.4}$$

$$F(W) \approx F(W^{(k)}) + \langle \partial F(W^{(k)}), W - W^{(k)} \rangle + \frac{\mu_k}{2} ||W - W^{(k)}||_F^2 \tag{5.5}$$

Taking Eq.(5.5) into Eq.(5.1), we have

$$W^{(k+1)} = \arg\min_W F(W^{(k)}) + \langle \partial F(W^{(k)}), W - W^{(k)} \rangle + \frac{\mu_k}{2} ||W - W^{(k)}||_F^2 + \lambda ||W||_\star$$

The above problem admits a closed form solution by using SVD thresholding operator to $M = W^{(k)} - \frac{1}{\mu_k} \partial F(W^{(k)})$.

Experiments

To evaluate the proposed LRR model on the manifold of SPD matrices, we apply it to both clean and corrupted image datasets for image classification and image segmentation.

Table: Input Feature Comparisons among Baselines

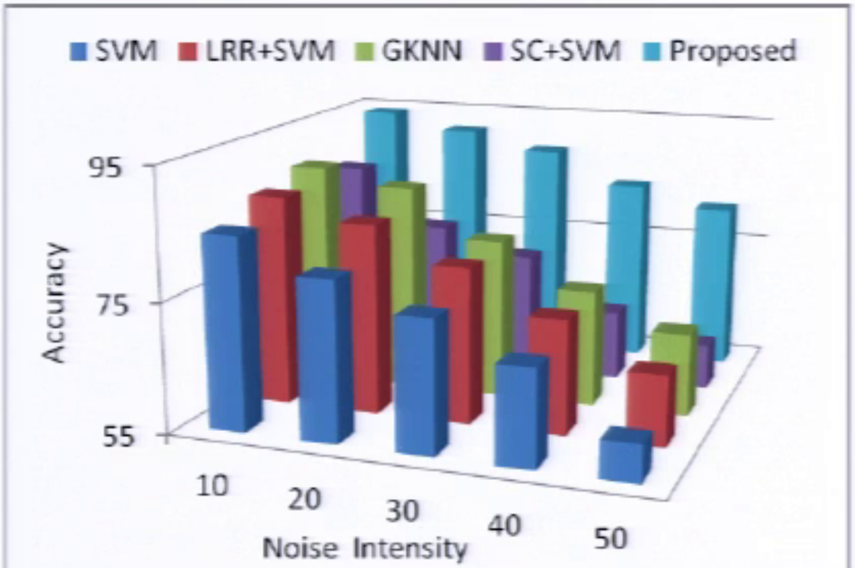
Baseline Methods	Input Features for Classifcation/Clustering
SVM/Ncut	Vectoried SPD matrices
LRR+SVM/Ncut	Low rank features on the Euclidean space
GKNN/RNcut	Original SPD matrix features
SC+SVM/our new method	Low rank features on the Manifold of SPD matrices

Performance for Image Classification

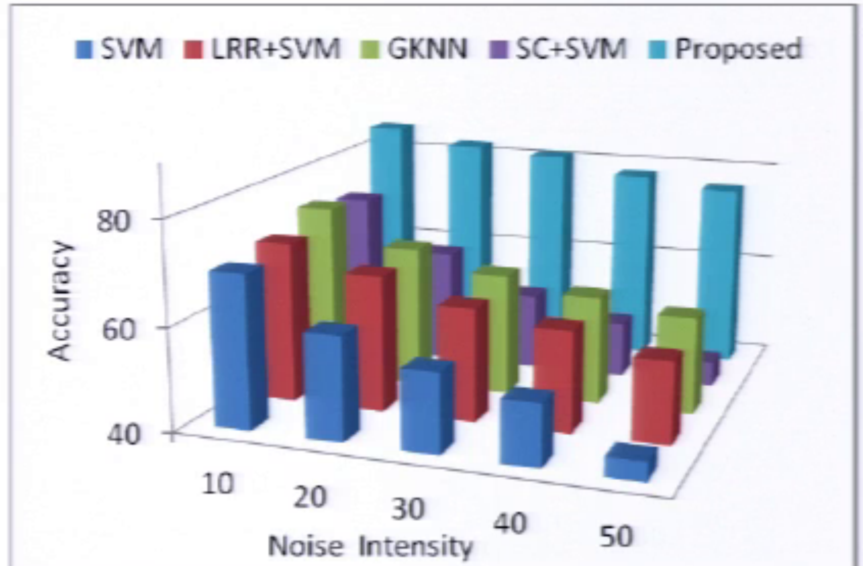
Table: Classification Accuracy comparisons on Clean data sets

	Brodatz		IDIAP		
Class	16	32	5	10	15
SVM	93.36	88.67	76.54	72.75	68.56
LRR+SVM	95.08	90.99	80.17	75.81	72.89
GKNN	95.7	92.11	82.69	77.32	80.36
SC+SVM	99.37	95.43	87.71	82.69	83.78
Proposed	99.89	97.12	90.38	87.78	87.33

Noise Robustness of the Proposed Model



(a) 16-class Brodatz



(b) 15-class IDIAP

Figure: Classification Accuracy comparisons on Noisy data sets

Performance for Image Segmentation

Table: Image segmentation accuracy on the Automatic Photo Pop-up dataset

Dataset	Ncut	LRR+Ncut	RNcut	Proposed
beach04	76.75	80.55	81.67	84.96
roads03	78.66	82.51	83.68	87.99
beach01	80.56	84.33	86.56	89.34
build05	75.44	78.34	80.54	83.55

Conclusions

- We propose a novel LRR model on the manifold of SPD matrices, in which we exploit the intrinsic property of SPD matrices manifold in the Riemannian geometric context.
- Compared with the existing Euclidean LRR algorithms, the loss of the global linear structure is compensated by the local linear structures given by the tangent spaces of the manifold.
- Further, we derive a easily solvable optimization problem, which incorporates the structured embedding mapping into the LRR model.
- Our experiments demonstrate that our proposed method is efficient and robust to the noise, and produces superior results compared to other state-of-art methods for classification and segmentation applications on several computer vision datasets.

Thanks! Questions?