APPROXIMATE INFINITE-DIMENSIONAL REGION COVARIANCE DESCRIPTORS FOR IMAGE CLASSIFICATION

Masoud Faraki, Mehrtash T. Harandi, Fatih Porikli

College of Engineering and Computer Science, Australian National University, Australia
NICTA, Canberra Research Laboratory, Australia
Email: {masoud.faraki,mehrtash.harandi,fatih.porikli}@nicta.com.au

ABSTRACT

We introduce methods to estimate infinite-dimensional Region Covariance Descriptors (RCovDs) by exploiting two feature mappings, namely random Fourier features and the Nyström method. In general, infinite-dimensional RCovDs offer better discriminatory power over their low-dimensional counterparts. However, the underlying Riemannian structure, i.e., the manifold of Symmetric Positive Definite (SPD) matrices, is out of reach to great extent for infinite-dimensional RCovDs. To overcome this difficulty, we propose to approximate the infinite-dimensional RCovDs by making use of the aforementioned explicit mappings. We will empirically show that the proposed finite-dimensional approximations of infinite-dimensional RCovDs consistently outperform the low-dimensional RCovDs for image classification task, while enjoying the Riemannian structure of the SPD manifolds. Moreover, our methods achieve the state-of-the-art performance on three different image classification tasks.

Index Terms— Region Covariance Descriptor, Reproducing Kernel Hilbert Space, Riemannian Geometry

1. INTRODUCTION

In this paper, we propose methods to approximate the recently introduced infinite-dimensional Region Covariance Descriptors (RCovDs) [1]. The motivation here stems from the fact that the Riemannian geometry -which is essential in analyzing RCovDs- does not apply verbatim to the infinite-dimensional case. Hence, by approximating the infinite-dimensional RCovDs with finite-dimensional ones, one could seamlessly exploit the rich geometry of RCovDs and tools developed upon that to do the inference.

RCovDs [2] are robust and fairly novel image descriptors that encode the second order statistics of features. One could think of it as capturing the relative correlation of features along their powers as a mean for representation. In computer vision community, RCovDs have been successfully employed to address various visual tasks. Notable examples include pedestrian detection [3], texture categorization [4], human epithelial cell classification [5], and DTI analysis [6].

Despite their attractiveness and success, RCovDs are SPD matrices and naturally lie on a connected Riemannian manifold. Consequently, Euclidean geometry is not appropriate to analyze them as shown in several recent studies [2, 6, 7, 8, 4].

In an attempt to encode more information in RCovDs, we have recently introduced infinite-dimensional RCovDs [1]. To this end, a mapping from the low-dimensional Euclidean space to a Reproducing Kernel Hilbert Space (RKHS), i.e., $\phi : \mathbb{R}^d \rightarrow \mathcal{H}$, is used along the kernel trick to compute several forms of Bregman divergences between infinite-dimensional RCovDs in $\mathcal{H}$. In practice, infinite-dimensional RCovDs are rank deficient. This is because a valid d-dimensional RCovD requires more than d independent observations which translates into the impractical situation of having endless observations for the infinite-dimensional RCovDs. This difficulty while partly resolved through regularization deprives us from exploiting the geometry of the space. More specifically, tangent spaces, exponential and logarithm maps, and geodesics are out of reach to our best knowledge.

In this paper, we overcome the aforementioned issue by introducing two methods to approximate infinite-dimensional RCovDs by finite-dimensional ones. To this end, we use random Fourier features [9] and the Nyström method [10] to learn a mapping $z : \mathbb{R}^d \rightarrow \mathbb{R}^D, d \leq D$ such that $\langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle_{\mathcal{H}} \simeq z(\boldsymbol{x}_i)^T z(\boldsymbol{x}_j)$. Having the mapping $z(\cdot)$ at our disposal, we approximate the infinite-dimensional RCovDs with $D \times D$ SPD matrices and take advantage of the Riemannian geometry of $\mathcal{S}_{++}^D$ [1] to analyze the resulting RCovDs. We will show that both methods constantly outperform the low-dimensional RCovDs and achieve the state-of-the-art performance on three challenging image classification tasks, namely material categorization, virus cell identification, and scene classification. Moreover, our experiment shows that the RCovDs in the learned space could even outperform the infinite-dimensional ones. This is of course inline with findings in [11, 12, 13].

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the ARC through the ICT Centre of Excellence program.

[1] The manifold of $D \times D$ SPD matrices.

2. RELATED WORK

We start this section by formally defining the region covariance descriptors [2]. Let $\boldsymbol{X} = \left[\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n\right], \boldsymbol{x}_i \in \mathbb{R}^d$ be a $d \times n$ matrix of n observations (extracted from an image or a video). The RCovD $\boldsymbol{C} \in \mathcal{S}_{++}^d$ as its name implies is defined as

$$\boldsymbol{C} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T = \boldsymbol{X}\boldsymbol{J}\boldsymbol{J}^T\boldsymbol{X}^T, \quad (1)$$

where $\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$ is the sample mean of the observations, $\boldsymbol{J} = n^{-\frac{3}{2}}(n\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^T)$, and $\mathbf{1}_n$ is a column vector of n ones.

Based on Eq. (1), an RCovD $\boldsymbol{C}_X$ in an RKHS $\mathcal{H}$ with dimensionality $|\mathcal{H}|$ can be defined as

$$\boldsymbol{C}_X = \Phi_{\boldsymbol{X}}\boldsymbol{J}\boldsymbol{J}^T\Phi_{\boldsymbol{X}}^T, \quad (2)$$

where $\Phi_{\boldsymbol{X}} = \left[\phi(\boldsymbol{x}_1)|\phi(\boldsymbol{x}_2)|\cdots|\phi(\boldsymbol{x}_n)\right]$ and $\phi : \mathbb{R}^d \to \mathcal{H}$ is the implicit mapping to $\mathcal{H}$.

While embeddings into an RKHS seems preferable in many applications, the applicability of infinite-dimensional RCovDs is limited. This is evident by considering the situation where the dimensionality $|\mathcal{H}|$ approaches ∞, which leads to $\boldsymbol{C}_X$ being semi-definite. As a consequence, $\boldsymbol{C}_X$ is on the boundary of the positive cone and at infinite distance form SPD matrices.

In the following two sections, we will show how an infinite-dimensional $\boldsymbol{C}_X$ can be approximated by a finite $D \times D$ one. But before delving into that, we establish some notations and definitions that will be used in the subsequent sections.

Definition 1 (Real-valued Positive Definite Kernels) *Let $\mathcal{X}$ be a nonempty set. A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite (**pd**) kernel on $\mathcal{X}$ if and only if $\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) > 0$ for any $n \in \mathbb{N}$, $x_i \in \mathcal{X}$ and non-zero vector $\boldsymbol{c} = (c_1, c_2, \cdots, c_n)^T \in \mathbb{R}^n$.*

According to Mercer's theorem, for any pd kernel $k(\cdot,\cdot)$, there exists a mapping $\phi : \mathcal{X} \to \mathcal{H}$ such that: $\forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}, k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle\phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j)\rangle_{\mathcal{H}}$.

Our main interest in this paper is the Riemannian manifold of $d \times d$ SPD matrices, i.e., $\mathcal{S}_{++}^d$. We will use $T_{\boldsymbol{P}}\mathcal{M}$ to show the tangent space of manifold $\mathcal{M}$ at point $\boldsymbol{P}$. For the $\mathcal{S}_{++}^d$, the tangent space is the space of symmetric matrices and the *logarithm map* $\log_{\boldsymbol{P}}(\cdot) : \mathcal{M} \to T_{\boldsymbol{P}}\mathcal{M}$ is identified by the principal matrix logarithm [14]. The Riemannian structure induced by the Affine Invariant Riemannian Metric (AIRM) [6] is considered the correct way of analyzing SPD matrices. The geodesic distance between points $\boldsymbol{C}_1$ and $\boldsymbol{C}_2 \in \mathcal{S}_{++}^d$ based on AIRM is

$$\delta_R(\boldsymbol{C}_1, \boldsymbol{C}_2) = \|\log(\boldsymbol{C}_1^{-1/2}\boldsymbol{C}_2\boldsymbol{C}_1^{-1/2})\|_F\ . \quad (3)$$

3. RANDOM FOURIER FEATURES

We start this section by providing a brief description of the method of random Fourier features for approximating $\phi(\cdot)$. Since in our experiments in § 5, we will only use RBF kernel, we limit the discussion here to this special kernel. The signature of other important kernels can be found in [9, 15].

According to the Bochner theorem [16], a shift-invariant kernel[2] such as RBF kernel can be can be obtained by the following Fourier integral

$$k(\boldsymbol{x}_i - \boldsymbol{x}_j) = \int_{\mathbb{R}^d} p(\omega) e^{j\omega^T\boldsymbol{x}_i} e^{-j\omega^T\boldsymbol{x}_j} d\omega. \quad (4)$$

In other words, $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k(\boldsymbol{x}_i - \boldsymbol{x}_j)$ is the expected value of $\zeta_\omega(\boldsymbol{x}_i)\zeta_\omega^*(\boldsymbol{x}_j)$ according to the distribution $p(\omega)$ where $\zeta_\omega(\boldsymbol{x}) = e^{j\omega^T\boldsymbol{x}}$. As shown in [9], the function $z_F(\boldsymbol{x}) = \sqrt{2}\cos(\omega^T\boldsymbol{x} + b)$ satisfies the aforementioned criterion for real kernels, i.e., $E[z_F(\boldsymbol{x}_i)z_F(\boldsymbol{x}_j)] = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ with ω and b being random variables drawn from $p(\omega)$ and $[0, 2\pi]$, respectively. For the RBF kernel $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|k(\boldsymbol{x}_i - \boldsymbol{x}_j)\|^2/2\sigma^2)$, $p(\omega) = \mathcal{N}(0, \sigma^{-2}\mathbf{I}_{d\times d})$ [9].

As such, let $\omega_1, \omega_2, \cdots, \omega_D$, $\omega_i \in \mathbb{R}^d$, be i.i.d samples drawn form the normal distribution $\mathcal{N}(0, \sigma^{-2}\mathbf{I}_{d\times d})$ and $b_1, b_2, \cdots, b_D$ be samples uniformly drawn from $[0, 2\pi]$. Then, the D dimensional estimation of $\phi(\boldsymbol{x})$ is given by

$$z_F(\boldsymbol{x}) = \sqrt{\frac{2}{D}}\Big[\cos(\omega_1^T\boldsymbol{x} + b_1), \cdots, \cos(\omega_D^T\boldsymbol{x} + b_D)\Big]. \quad (5)$$

Having the mapping $z_F : \mathbb{R}^d \to \mathbb{R}^D$ at our disposal, our first estimation of an infinite-dimensional RCovD can be obtained as

$$\hat{\boldsymbol{C}}_X = \Phi_{\boldsymbol{X}}\boldsymbol{J}\boldsymbol{J}^T\Phi_{\boldsymbol{X}}^T, \quad (6)$$

where $\Phi_{\boldsymbol{X}} = \left[z_F(\boldsymbol{x}_1)|z_F(\boldsymbol{x}_2)|\cdots|z_F(\boldsymbol{x}_n)\right]$.

Algorithm 1 outlines the details of computing RCovDs using random Fourier features for the RBF kernel.

4. NYSTRÖM METHOD

While in § 3, an approximation to the embedding function $\phi(\cdot)$ was provided, we note that not only an arbitrary kernel $k(\cdot,\cdot)$ may not satisfy the Bochner theorem (e.g., if it is not shift-invariant), but even if it is, it may not be possible to obtain $p(\omega)$ analytically. To alleviate this limitation, we propose a data-dependent estimation of the RKHS $\mathcal{H}$ using the Nyström method [10].

Given $\mathcal{D} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$ a collection of M training examples[3], a rank D approximation of $\boldsymbol{K} = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{M\times M}$ can be written as $\boldsymbol{Z}^T\boldsymbol{Z}$. Here, $\boldsymbol{Z}_{D\times M} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{V}$ with $\boldsymbol{\Sigma}$ and

[2] A kernel function is shift invariant if $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k(\boldsymbol{x}_i - \boldsymbol{x}_j)$.

[3] Observations extracted from training images in our case.

Algorithm 1 Approximate infinite-dimensional RCovD using random Fourier features

Input:

- $\boldsymbol{X} = \big[\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n\big]$, $\boldsymbol{x}_i \in \mathbb{R}^d$, matrix of n feature vectors
- σ^2, scale of the RBF kernel
- D, target dimensionality

Output:

- $\hat{\boldsymbol{C}}_X \in \mathcal{S}_{++}^D$, approximate infinite-dimensional RCovD

1: $\{\omega_i\}_{i=1}^D \leftarrow$ i.i.d samples drawn from $\mathcal{N}(0, \sigma^{-2}\mathbf{I}_{d\times d})$.
2: $\{b_i\}_{i=1}^D \leftarrow$ uniform samples drawn from $[0, 2\pi]$.
3: **for** $j = 1 \rightarrow n$ **do**
4: Compute $z_F(\boldsymbol{x}_j)$ using Eq. (5).
5: **end for**
6: Compute $\hat{\boldsymbol{C}}_X$ using Eq. (6)

Algorithm 2 Approximate infinite-dimensional RCovD using the Nyström method

Input:

- $\boldsymbol{X} = \big[\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n\big]$, $\boldsymbol{x}_i \in \mathbb{R}^d$, matrix of n feature vectors
- $\mathcal{D} = \{\boldsymbol{x}_i\}_{i=1}^M, \boldsymbol{x}_i \in \mathbb{R}^d$, a collection of training examples
- D, target dimensionality

Output:

- $\hat{\boldsymbol{C}}_X \in \mathcal{S}_{++}^D$, approximate infinite-dimensional RCovD

1: Compute the kernel matrix $\boldsymbol{K} = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{M\times M}$.
2: $\boldsymbol{\Sigma} \leftarrow$ diagonal matrix of top D eigenvalues of $\boldsymbol{K}$.
3: $\boldsymbol{V} \leftarrow$ associated eigenvectors of $\boldsymbol{\Sigma}$.
4: **for** $j = 1 \rightarrow n$ **do**
5: Compute $z_N(\boldsymbol{x}_j)$ using Equation 7.
6: **end for**
7: Compute $\hat{\boldsymbol{C}}_X$ using Equation 8.

$\boldsymbol{V}$ being the top D eigenvalues and corresponding eigenvectors of $\boldsymbol{K}$. Based on this low-rank approximation, one can obtain a D-dimensional vector representation of the space $\boldsymbol{K}$ as

$$z_N(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\Big(k(\boldsymbol{x}, \boldsymbol{x}_1), \cdots, k(\boldsymbol{x}, \boldsymbol{x}_M)\Big)^T. \quad (7)$$

Given $\boldsymbol{X} = \big[\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n\big]$, a set of n observations, the corresponding RKHS region covariance descriptor estimation using the Nyström method is obtained as

$$\hat{\boldsymbol{C}}_X = \Phi_{\boldsymbol{X}}\boldsymbol{J}\boldsymbol{J}^T\Phi_{\boldsymbol{X}}^T, \quad (8)$$

where $\Phi_{\boldsymbol{X}} = \big[z_N(\boldsymbol{x}_1)|z_N(\boldsymbol{x}_2)|\cdots|z_N(\boldsymbol{x}_n)\big]$.

Algorithm 2 summarizes the discussion about estimating RCovDs using the Nyström method in one pseudo-code.

5. EXPERIMENTS

In this section, we evaluate the proposed approximate infinite-dimensional RCovDs on three different classification tasks, namely material categorization, virus cell identification, and scene classification. For benchmarking, we compare the accuracy of the Nearest Neighbor (NN) classifier on low-dimensional manifold against NN in higher-dimensional manifolds obtained by random Fourier features or the Nyström method.

Beside NN classifier, we will evaluate the performance of the state-of-the-art method of Covariance Discriminant Learning (CDL) [17] for low and high-dimensional SPD manifolds. The CDL technique utilizes the identity tangent space of the SPD manifold to perform kernel Partial Least Squares (kPLS) [18]. Partial Least Squares (PLS) [19] can be understood as a dimensionality reduction technique that models relations between two sets of variables through a latent space. In the context of classification, PLS and its kernelized version can be used to model the relations between feature vectors and their representative classes.

The different algorithms evaluated in our experiments are referred to as

- **NN**: AIRM based NN classifier on low-dimensional RCovDs.
- **NN$_\mathbf{F}$**: AIRM based NN classifier on approximate infinite-dimensional RCovDs obtained by random Fourier features.
- **NN$_\mathbf{N}$**: AIRM based NN classifier on approximate infinite-dimensional RCovDs obtained by the Nyström method.
- **CDL**: CDL on low-dimensional RCovDs.
- **CDL$_\mathbf{F}$**: CDL on approximate infinite-dimensional RCovDs obtained by random Fourier features.
- **CDL$_\mathbf{N}$**: CDL on approximate infinite-dimensional RCovDs obtained by the Nyström method.

In what follows, we first elaborate on how rich RCovDs can be obtained for each task. This is followed by in-depth discussions on the performance of approximate infinite-dimensional RCovDs obtained through the processes described in § 3 and § 4, respectively.

5.1. Material Categorization

Material categorization is the task of classifying materials from their appearance in single images taken under unknown viewpoint and illumination conditions. For this experiment, we have used the UIUC material classification dataset [20] which contains 18 classes of complex material categories "taken in the wild" (see Fig. 1 for sample images). The images were mainly selected to have various geometric fine-scale details. We split the database into training and test sets by randomly assigning half of the images of each class to the training set and using the rest as test data. The process of random splitting was repeated 10 times and the average recognition accuracies along standard deviations will be reported here.

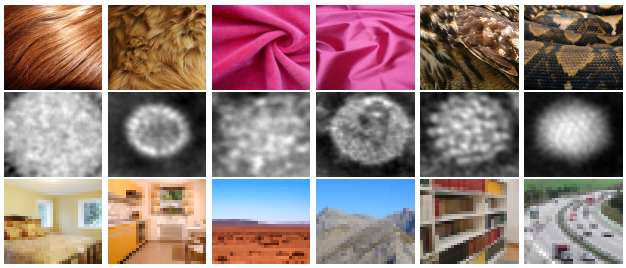

Fig. 1: Sample images for datasets used in this work. Top: UIUC [20], Middle: Virus [22], Bottom: TinyGraz03 [23].

To generate RCovDs, a feature vector is assigned to each pixel at position (x, y) in an image I by

$$F_{(x,y)} = \left[I_R(x,y), I_G(x,y), I_B(x,y), \left|\frac{\partial I}{\partial x}\right|, \left|\frac{\partial I}{\partial y}\right|, \left|\frac{\partial^2 I}{\partial x^2}\right|, \left|\frac{\partial^2 I}{\partial y^2}\right|, |G_{(0,0)}(x,y)|, \cdots, |G_{(u,v)}(x,y)|\right], \quad (9)$$

where $I_c(x,y), c \in \{R,G,B\}$, denotes color information, the next four entries are the magnitude of intensity gradients and the magnitude of Laplacians along x and y directions, and $G_{(u,v)}(x,y)$ is the response of a 2D Gabor wavelet [21] centered at (x,y) with orientation u and scale v. We extracted Gabor wavelets at four orientations and three scales. Therefore, each pixel is described by a 19 dimensional feature vector (i.e., 3 color, 4 gradients, and 12 Gabor features).

Table 1 shows the recognition accuracies for the studied methods. The correct classification rates obtained by NN clearly show that the proposed RCovDs are more discriminative than their low-dimensional counterparts. We also note that $\mathbf{NN_F}$ and $\mathbf{NN_N}$ achieve comparable performances to the more involved CDL in low-dimensional manifold.

The state-of-the-art performance on this dataset is 43.5% reported by [20]. CDL on the proposed RCovDs (both random Fourier features and Nyström) outperforms the state-of-the-art performance by at least 2.8% percentage points.

5.2. Virus Classification

We performed an experiment to classify cell images using the Virus dataset [22]. The dataset contains 1500 images of 15 different classes (100 samples per class). The images are formed from Transmission Electron Microscopy technique and re-sampled to 41×41 pixel grayscale image (see Fig. 1 for examples). Here, RCovDs are obtained using the features described in Eq. (9) with one modification. For this task, we used Gabor wavelets at four orientations and five scales.

Our empirical results are reported in Table 1. The average correct recognition rate with both $\mathbf{CDL_F}$ and $\mathbf{CDL_N}$ is superior to the state-of-the-art performance of 81.2% reported in [1] using infinite-dimensional RCovDs. We conjecture that computing the RCovDs with both random Fourier features and the Nyström method reveals the nonlinear patterns in data (as also evidenced in [11]). This is emphasized by the Riemannian structure of $\mathcal{S}_{++}^{D}$ (as CDL requires its tangent space) which is not available for the infinite-dimensional RCovDs.

Table 1: Recognition accuracies for the UIUC [20], Virus [22], and TinyGraz03 [23] datasets.

Method	UIUC	Virus	TinyGraz03
NN	26.5% ± 3.7	58.8% ± 5.4	34%
$\mathbf{NN_F}$	35.9% ± 3.0	67.1% ± 4.2	42%
$\mathbf{NN_N}$	35.6% ± 2.7	69.5% ± 4.8	44%
CDL	36.3% ± 2.0	75.5% ± 2.5	41%
$\mathbf{CDL_F}$	**47.4% ± 3.1**	**82.5% ± 2.9**	55%
$\mathbf{CDL_N}$	46.3% ± 2.6	81.4% ± 3.1	**57%**

5.3. Scene Classification

For the last experiment, we considered the task of scene classification using TinyGraz03 dataset [23]. The dataset contains 1148 indoor and outdoor images (see Fig. 1 for examples) with a spatial resolution of 32×32 pixels.The images are presented in 20 classes with at least 40 samples per class. This dataset is quite diverse, with scene categories being captured from various viewpoints and under various lighting conditions. We used the recommended train/test split provided by the authors. The correct recognition rate achieved by humans on this dataset is 30% [23].

The RCovDs for this task were obtained using the first 7 features in Eq. (9) (i.e., 3 color and 4 image gradients). Table 1 indicates that computing RCovDs using random Fourier features and the Nyström method offers notable enhancement in term of discriminatory power over the original RCovDs. We also note that $\mathbf{NN_F}$ and $\mathbf{NN_N}$ outperform the more involved $\mathbf{CDL}$.

The state-of-the-art recognition accuracy on this dataset is reported to be 46% [23]. Interestingly, $\mathbf{CDL_F}$ and $\mathbf{CDL_N}$ significantly outperform the state-of-the-art method (more than 9 percentage points) and human performance (more than 25 percentage points).

6. CONCLUSIONS AND FUTURE WORK

We have made use of random Fourier feature and the Nyström method to compute two approximations to infinite-dimensional RCovDs. Our experimental evaluation has demonstrated that the proposed RCovDs significantly outperform the low-dimensional ones on image classification task. More importantly, our RCovDs provide a framework in which the well-known Riemannian geometry of the SPD matrices can be taken into account. In the future, we intend to explore how the proposed approach can be extended to other types of Riemannian manifolds, such as Grassmannian manifolds.

7. REFERENCES

[1] Mehrtash Harandi, Mathieu Salzmann, and Fatih Porikli, "Bregman divergences for infinite dimensional covariance matrices," *arXiv preprint arXiv:1403.4334*, 2014.

[2] Oncel Tuzel, Fatih Porikli, and Peter Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. European Conference on Computer Vision (ECCV)*, 2006, pp. 589–600.

[3] Oncel Tuzel, Fatih Porikli, and Peter Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 30, no. 10, pp. 1713–1727, 2008.

[4] Mehrtash T Harandi, Conrad Sanderson, Richard Hartley, and Brian C Lovell, "Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 216–229. Springer, 2012.

[5] Masoud Faraki, Mehrtash T Harandi, Arnold Wiliem, and Brian C Lovell, "Fisher tensors for classifying human epithelial cells," *Pattern Recognition (PR)*, vol. 47, no. 7, pp. 2348–2359, 2014.

[6] Xavier Pennec, Pierre Fillard, and Nicholas Ayache, "A riemannian framework for tensor computing," *Int. Journal of Computer Vision (IJCV)*, vol. 66, no. 1, pp. 41–66, 2006.

[7] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 1, pp. 328–347, 2007.

[8] Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi, "Kernel methods on the riemannian manifold of symmetric positive definite matrices," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 73–80.

[9] Ali Rahimi and Benjamin Recht, "Random features for large-scale kernel machines," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2007, pp. 1177–1184.

[10] Christopher TH Baker, *The numerical treatment of integral equations*, Clarendon press, 1977.

[11] David Lopez-Paz, Suvrit Sra, Alex Smola, Zoubin Ghahramani, and Bernhard Schölkopf, "Randomized nonlinear component analysis," *arXiv preprint arXiv:1402.0119*, 2014.

[12] Ali Rahimi and Benjamin Recht, "Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2009, pp. 1313–1320.

[13] Quoc Le, Tamás Sarlós, and Alex Smola, "Fastfoodapproximating kernel expansions in loglinear time," in *Proc. Int. Conference on Machine Learning (ICML)*, 2013.

[14] Rajendra Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[15] Andrea Vedaldi and Andrew Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 3, pp. 480–492, 2012.

[16] Walter Rudin, *Fourier analysis on groups*, John Wiley & Sons, 2011.

[17] Ruiping Wang, Huimin Guo, Larry S Davis, and Qionghai Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2496–2503.

[18] Roman Rosipal and Leonard J Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *Journal of Machine Learning Research (JMLR)*, vol. 2, pp. 97–123, 2002.

[19] Roman Rosipal and Nicole Krämer, "Overview and recent advances in partial least squares," in *Subspace, Latent Structure and Feature Selection*, pp. 34–51. Springer, 2006.

[20] Zicheng Liao, Jason Rock, Yang Wang, and David Forsyth, "Non-parametric filtering for geometric detail extraction and material representation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 963–970.

[21] Tai Sing Lee, "Image representation using 2d Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 18, pp. 959–971, 1996.

[22] Gustaf Kylberg, Mats Uppström, and Ida-Maria Sintorn, "Virus texture analysis using local binary patterns and radial density profiles," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 573–580. Springer, 2011.

[23] Andreas Wendel and Axel Pinz, "Scene categorization from tiny images," in *31st Annual Workshop of the Austrian Association for Pattern Recognition*, 2007, pp. 49–56.