Kernel Analysis over Riemannian Manifolds for Visual Recognition of Actions, Pedestrians and Textures

Mehrtash T. Harandi, Conrad Sanderson, Arnold Wiliem, 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 analysing Riemannian manifolds is to embed them in Euclidean spaces, with the embedding typically obtained by flattening the manifold via tangent spaces. This general approach is not free of drawbacks. For example, only distances between points to the tangent pole are equal to true geodesic distances. This is restrictive and may lead to inaccurate modelling. Instead of using tangent spaces, we propose embedding into the Reproducing Kernel Hilbert Space by introducing a Riemannian pseudo kernel. We furthermore propose to recast a locality preserving projection technique from Euclidean spaces to Riemannian manifolds, in order to demonstrate the benefits of the embedding. Experiments on several visual classification tasks (gesture recognition, person re-identification and texture classification) show that in comparison to tangent-based processing and state-of-the-art methods (such as tensor canonical correlation analysis), the proposed approach obtains considerable improvements in discrimination accuracy.

1. Introduction

Recently, non-Euclidean geometry, such as Riemannian manifolds, has opened new ways to interpret and analyse image as well as video data [14, 18, 19, 25, 26, 28]. The curious mind might ask what are the motivations and advantages of switching from the well-defined Euclidean spaces to curved, Riemannian spaces? A short answer to this question would be – *the features and visual models often do not lie on an Euclidean space*. In other words the underlying distance function on the space is not the usual Euclidean L_p norm. As such, Riemannian manifolds might be an appropriate way of inference in various regimes of visual computation, especially the identification paradigm.

In this paper we consider the space formed by non-singular covariance matrices, which are symmetric positive definite matrices. Such matrices form a connected Riemannian manifold, not an Euclidean space [28]. Covariance matrices as region descriptors were first introduced by Tuzel *et al*. [27] and since then have been employed successfully for object tracking [20], pedestrian detection [28], action recognition [11] and medical imaging [19].

Prior Work. Inference on Riemannian manifolds can be achieved by embedding the manifolds in higher dimensional Euclidean spaces, which can be considered as flattening the manifold. In the literature, the most popular choice for embedding the manifold is through considering tangent spaces [11, 20, 28]. Tuzel *et al*. [28] tackled the problem of pedestrian detection by designing a LogitBoost classifier [8] over Riemannian manifold spaces. Due to the curvature of the space, Tuzel et al. designed each weak classifier on an appropriate tangent space. As such, the inference on the manifold was made through several tangent spaces. Subbarao *et al*. [25] reformulated the mean shift algorithm [4] over non-linear manifolds. In particular they showed that the mean shift can be seen as an iterative approach that switches between manifold and tangent spaces. For action classification, Guo et al. [11] proposed to a sparse-based solution on Riemannian manifolds by mapping all the points on the manifold to the tangent space of the identity matrix.

Flattening the manifold through tangent spaces is not free of drawbacks. For example, only distances between points to the tangent pole are are equal to true geodesic distances. This is restrictive and may lead to inaccurate modelling. A recent alternate school of thought considers embedding Grassmann manifolds (a special case of Riemannian manifolds) into Reproducing Kernel Hilbert Spaces (RKHS) [24], through the use of dedicated Grassmann kernel functions [12, 14]. This in turn opens the door for employing many kernel-based machine learning algorithms [24].

Contributions. There are two main novelties in this work. Firstly, based on the Riemannian geodesic distance, we propose a Riemannian pseudo kernel. Unlike the kernels used in [12, 14], the proposed kernel is not restricted to any special class of Riemannian manifolds. Secondly, having a kernel at our disposal, we exploit RKHS theory to recast a locality preserving projection method [15] from Euclidean vector spaces to Riemannian manifolds. Lastly, we apply the proposed approach to 3 distinct visual classification tasks: recognition of actions, textures and pedestrians.

We continue the paper as follows. Section 2 provides a brief overview of Riemannian manifolds, which leads to the proposed Riemannian pseudo kernel in Section 3. In Section 4 we recast Euclidean locality preserving projection to Riemannian manifolds. In Section 5 we compare the performance of the proposed method with previous approaches on the abovementioned visual classification tasks. The main findings and possible future directions are summarised in Section 6.

2. Riemannian Geometry

In this section we briefly review Riemannian geometry, with a focus on the space of symmetric positive definite matrices. Formally, a manifold is a topological space which is locally similar to an Euclidean space [28]. Intuitively, we can think of a manifold as a continuous surface lying in a higher dimensional Euclidean space.

The tangent space, $T_{\boldsymbol{X}}$ at $\boldsymbol{X}$, is the plane tangent to the surface of the manifold at that point. The tangent space can be thought of as the set of allowable velocities for a point constrained to move on the manifold. The minimum length curve connecting two points on the manifold is called the geodesic, and the distance between two points $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by the length of this curve.

For a Riemannian manifold, geodesics (on the manifold) are related to the tangents in the tangent space. For each tangent $\Delta \in T_{\boldsymbol{X}}$, there exists a unique geodesic starting at $\boldsymbol{X}$ with initial velocity Δ. Two operators, namely the exponential $\exp_{\boldsymbol{X}}$ and logarithm maps $\log_{\boldsymbol{X}} = \exp_{\boldsymbol{X}}^{-1}$, are defined over the Riemannian manifolds to switch between manifold and tangent space at $\boldsymbol{X}$. More specifically, the exponential operator maps Δ to the point $\boldsymbol{Y}$ on the manifold. The property of the exponential map ensures that the length of Δ is equivalent to the geodesic distance between $\boldsymbol{X}$ and $\boldsymbol{Y}$. The logarithm map is the inverse of the exponential map and maps a point on the manifold to the tangent space $T_{\boldsymbol{X}}$. The exponential and logarithm operators vary as point $\boldsymbol{X}$ moves. These concepts are illustrated in Fig. 1.

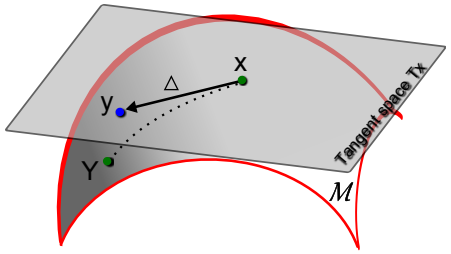

Figure 1. Illustration of the tangent space $T_{\boldsymbol{X}}$ at point $\boldsymbol{X}$ on a Riemannian manifold $\mathcal{M}$. A covariance matrix can be interpreted as point $\boldsymbol{X}$ in the space of symmetric positive definite matrices. The tangent vector Δ can be obtained through the logarithm mapping, ie. $\Delta = \log_{\boldsymbol{X}}(\boldsymbol{Y})$. Every tangent vector in $T_{\boldsymbol{X}}$ can be mapped back to the manifold through the exponential map, ie. $\exp_{\boldsymbol{X}}(\Delta) = \boldsymbol{Y}$. The dotted line shows the geodesic starting at $\boldsymbol{X}$ and ending at $\boldsymbol{Y}$.

Symmetric positive definite matrices with size $d \times d$, eg. non-singular covariance matrices, can be formulated as a connected Riemannian manifold (Sym_d^+). For Sym_d^+ the exponential and logarithm maps are defined as:

$$\exp_{\boldsymbol{X}}(\boldsymbol{y}) = \boldsymbol{X}^{\frac{1}{2}} \exp\left(\boldsymbol{X}^{-\frac{1}{2}} \boldsymbol{y} \boldsymbol{X}^{-\frac{1}{2}}\right) \boldsymbol{X}^{\frac{1}{2}} \quad (1)$$

$$\log_{\boldsymbol{X}}(\boldsymbol{Y}) = \boldsymbol{X}^{\frac{1}{2}} \log\left(\boldsymbol{X}^{-\frac{1}{2}} \boldsymbol{Y} \boldsymbol{X}^{-\frac{1}{2}}\right) \boldsymbol{X}^{\frac{1}{2}} \quad (2)$$

In (1) and (2), $\exp(\cdot)$ and $\log(\cdot)$ are matrix exponential and logarithm operators, respectively. For symmetric positive definite matrices they can be computed through Singular Value Decomposition (SVD). More specifically, let $\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{U}^T$ be the SVD of the symmetric matrix $\boldsymbol{X}$, then

$$\exp(\boldsymbol{X}) = \boldsymbol{U} \exp(\Sigma) \boldsymbol{U}^T \quad (3)$$

$$\log(\boldsymbol{X}) = \boldsymbol{U} \log(\Sigma) \boldsymbol{U}^T \quad (4)$$

In the above equations, $\exp(\Sigma)$ and $\log(\Sigma)$ are two diagonal matrices where the diagonal elements are respectively equivalent to the exponential or logarithms of the diagonal elements of matrix Σ.

3. Riemannian Kernel

By considering the geodesic distance between Riemannian points, we propose the following pseudo kernel:

$$k_{\mathbb{R}}(\boldsymbol{X}, \boldsymbol{Y}) = \exp\{-\sigma^{-1} d_G(\boldsymbol{X}, \boldsymbol{Y})\} \quad (5)$$

where $d_G(\boldsymbol{X}, \boldsymbol{Y}) = \text{trace}\left\{\log^2\left(\boldsymbol{X}^{-\frac{1}{2}} \boldsymbol{Y} \boldsymbol{X}^{-\frac{1}{2}}\right)\right\}$ for Sym_d^+.

Under certain conditions the proposed kernel becomes a true kernel (ie. a positive definite kernel function on $\mathcal{M}$). Specifically, the kernel matrix $\mathbb{K} = [k_{ij}]; k_{ij} = k_{\mathbb{R}}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ is positive definite iff $\boldsymbol{V}^T \mathbb{K} \boldsymbol{V} > 0, \forall \boldsymbol{V} \in \mathbb{R}^n$. Expanding $\boldsymbol{V}^T \mathbb{K} \boldsymbol{V}$ yields:

$$\begin{aligned} \boldsymbol{V}^T \mathbb{K} \boldsymbol{V} &= \left(\sum_{i=1}^n v_i\right)^2 - 2\sum_{i=1}^n \sum_{j \neq i} v_i v_j + 2\sum_{i=1}^n \sum_{j \neq i} v_i v_j k_{i,j} \\ &= \left(\sum_{i=1}^n v_i\right)^2 + 2\sum_{i=1}^n \sum_{j \neq i} v_i v_j (k_{i,j} - 1) \end{aligned} \quad (6)$$

Note that $k_{i,j} \in [0, 1]$. For values of v_i and v_j where $\min(v_i v_j (k_{i,j} - 1)) = -v_i v_j$ holds, we obtain:

$$\min\left(\boldsymbol{V}^T \mathbb{K} \boldsymbol{V}\right) = \left(\sum_{i=1}^n v_i\right)^2 - 2\sum_{i=1}^n \sum_{j \neq i} v_i v_j \quad (7)$$

As the right-hand side of Eqn. (7) is positive for $v_i \neq 0$, $\mathbb{K}$ would be a positive-definite matrix.

While the proposed pseudo kernel is not guaranteed to always be a positive definite function, experiments in Section 5 indicate that it can nevertheless still be quite useful. We note that it is possible to convert pseudo kernels into true kernels, as discussed in [3].

4. Riemannian Locality Preserving Projection

Given an affinity graph in a vector space, the purpose of locality preserving projections is to minimise an objective function that incurs a heavy penalty if neighbouring points in the original space are mapped far apart in the transformed space [15]. This problem can be solved through a generalised eigen-analysis framework. In the following text, we formulate the locality preserving projections over Riemannian manifolds. We call the resulting algorithm Riemannian Locality Preserving Projection (RLPP).

Given N points $\mathbb{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N\}$ from the underlying Riemannian manifold $\mathcal{M}$, the local geometrical structure of $\mathcal{M}$ can be modelled by building a similarity graph $\boldsymbol{W}$. The simplest form of $\boldsymbol{W}$ is a binary graph obtained based on the nearest neighbour properties of Riemannian points:

- ϵ-neighbourhoods. Two nodes are connected if the geodesic distance between them is less than a threshold.
- k nearest neighbours. Two nodes are connected by an edge if one node is among the k nearest neighbours of the other node.

We note that more complex affinity graphs can also be used to encode distances between points on Riemannian manifolds [24]. Our aim is to find a mapping from $\mathcal{M}$ to $\mathcal{M}'$, ie. $\alpha : \boldsymbol{X}_i \rightarrow \boldsymbol{Y}_i$, to preserve the local geometry of the manifold. A suitable transform would place the connected points of $\boldsymbol{W}$ as close as possible, while being flexible to some extent for the unconnected points of $\boldsymbol{W}$. Such a mapping can be described by optimising the following objective function:

$$f = \min \frac{1}{2} \sum\nolimits_{i,j} \left(\boldsymbol{Y}_i - \boldsymbol{Y}_j\right)^2 W(i,j) \tag{8}$$

Eqn. (8) punishes connected neighbours if they are mapped far away in $\mathcal{M}'$. Assume that points on the manifold are implicitly known and only a measure of similarity between them is available through a Riemannian kernel, denoted as $k_{ij} = \langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle$.

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

$$\boldsymbol{Y}_i = \left(\langle \boldsymbol{\alpha}_1, \boldsymbol{X}_i \rangle, \langle \boldsymbol{\alpha}_2, \boldsymbol{X}_i \rangle, \cdots, \langle \boldsymbol{\alpha}_r, \boldsymbol{X}_i \rangle\right)^T \tag{9}$$

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. (8) can be simplified to:

$$\begin{aligned} & \tfrac{1}{2} \textstyle\sum_{i,j} \left(\boldsymbol{Y}_i - \boldsymbol{Y}_j\right)^2 W(i,j) \\ = \; & \textstyle\sum_i \boldsymbol{A}_i^T \boldsymbol{K}_i \boldsymbol{K}_i^T \boldsymbol{A}_i^T W(i,i) - \sum_{i,j} \boldsymbol{A}_i^T \boldsymbol{K}_j \boldsymbol{K}_i^T \boldsymbol{A}_i^T W(i,j) \\ = \; & \mathbb{A}^T \mathbb{K} \boldsymbol{D} \mathbb{K}^T \mathbb{A} - \mathbb{A}^T \mathbb{K} \boldsymbol{W} \mathbb{K}^T \mathbb{A} \\ = \; & \mathbb{A}^T \mathbb{K} \boldsymbol{L} \mathbb{K}^T \mathbb{A} \end{aligned} \tag{10}$$

Algorithm 1. Pseudocode for training Riemannian Locality Preserving Projection (RLPP).

Input:

- Training set $\mathbb{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N\}$ from the underlying Riemannian manifold.
- The Riemannian heat kernel function k_{ij}, for measuring the similarity between two points on a Riemannian manifold.

Processing:

1. Compute the Gram matrix $[\mathbb{K}]_{ij}$ for all $\boldsymbol{X}_i, \boldsymbol{X}_j$
2. Compute the similarity graph, its degree and Laplacian matrices, $\boldsymbol{W}$, $\boldsymbol{D}$, and $\boldsymbol{L}$ respectively.
3. Solve the minimisation problem in Eqn. (11) by eigen decomposition to obtain $\mathbb{A}$. The r smallest eigenvectors of the Rayleigh quotient $\frac{\mathbb{K}\boldsymbol{D}\mathbb{K}^T}{\mathbb{K}\boldsymbol{L}\mathbb{K}^T}$ form $\mathbb{A}$.

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 an ascending manner according to their corresponding eigenvalues.

where $\mathbb{A} = [\boldsymbol{A}_1 | \boldsymbol{A}_2 | \cdots | \boldsymbol{A}_r]$, $\mathbb{K} = [\boldsymbol{K}_1 | \boldsymbol{K}_2 | \cdots | \boldsymbol{K}_N]$ and $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ is the Laplacian matrix. The minimum of (10) can be found by imposing the constraint $\mathbb{A}^T \mathbb{K} \boldsymbol{D} \mathbb{K}^T \mathbb{A} = 1$ [15, 30]. Hence we are interested in solving

$$\begin{aligned} & \arg\min_{\mathbb{A}} \;\; \mathbb{A}^T \mathbb{K} \boldsymbol{L} \mathbb{K}^T \mathbb{A} \\ & s.t. \quad \mathbb{A}^T \mathbb{K} \boldsymbol{D} \mathbb{K}^T \mathbb{A} = 1 \end{aligned} \tag{11}$$

The solution of (11) can be found through the following generalised eigenvalue problem:

$$\mathbb{K} \boldsymbol{L} \mathbb{K}^T \mathbb{A} = \lambda \mathbb{K} D \mathbb{K}^T \mathbb{A} \tag{12}$$

Algorithm 1 outlines the locality preserving projection on Riemannian manifolds. The algorithm uses the points on the Riemannian 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 preserves a measure of local similarity.

Upon acquiring the mapping $\mathbb{A}$, the matching problem over Riemannian manifolds is reduced to classification in vector spaces. More precisely, for any query sample $\boldsymbol{X}_q$, 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 $\boldsymbol{K}_q = \left(\langle \boldsymbol{X}_1, \boldsymbol{X}_q \rangle, \langle \boldsymbol{X}_2, \boldsymbol{X}_q \rangle, \cdots, \langle \boldsymbol{X}_N, \boldsymbol{X}_q \rangle\right)^T$. 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-Neighbours or Support Vector Machines [2] can be employed to label $\boldsymbol{X}_q$.

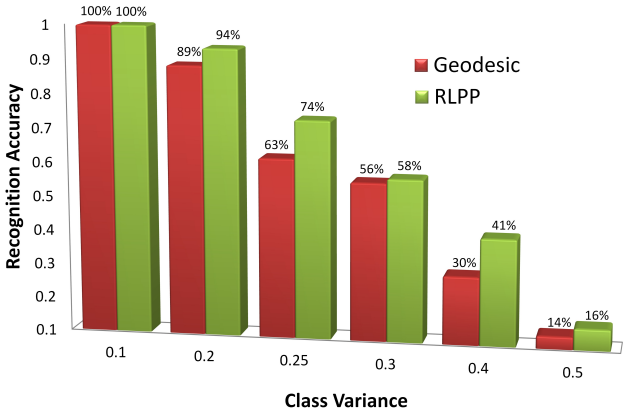

Figure 2. Comparison of the proposed RLPP approach with tangent-based analysis on synthetic data.

5. Experiments

We start this section by evaluating the performance of the proposed RLPP method[1] on synthetic data. We then compare and contrast RLPP to previous state-of-the-art methods on several classification tasks, including gesture recognition, texture classification and person re-identification.

5.1. Synthetic Data

For the synthetic data, we consider a multi-class classification problem over Sym_3^+. Since we are interested in contrasting tangent-based analysis with the proposed approach, we considered several classification problems on the identity tangent space (the space created by considering the identity matrix as the pole or centre of projection).

We randomly generated 16 classes over the identity tangent space where the samples in each class obey a normal distribution. Then all the generated samples were mapped back to the manifold using the exponential map. By fixing the mean of each class and increasing the class variance we created several classification problems with increasing difficulty.

Fig. 2 demonstrates that RLPP obtains superior performance when compared with tangent-based inference. We note that by increasing the class variance, samples of different classes are intertwined which leads to a decrease in recognition accuracy.

5.2. Gesture Recognition

For the hand-gesture recognition task, we used the Cambridge hand-gesture dataset [16] which consists of 900 image sequences of 9 gesture classes. Each class has 100 image sequences performed by 2 subjects, captured under 5 illuminations and 10 arbitrary motions. The 9 classes are defined by the 3 primitive hand shapes and 3 primitive motions. Each sequence was recorded at 30 fps with a resolution of 320×240, in front of a fixed camera. The gestures are roughly isolated in space and time. See Fig. 3 for examples. We follow the test protocol defined in [16], where sequences with normal illumination are considered for training while tests are performed on the remaining sequences.

The descriptor for a video sequence is obtained by computing the covariance matrix of frame descriptors. In a similar manner to [22], each frame descriptor is obtained by dividing the image into n_R rectangular regions and concatenating the descriptors from each region. There is no overlap between adjacent regions. Each region is further split into small (8×8) overlapping blocks. The amount of overlap between two adjacent blocks is n_p pixels. The region descriptor is simply the average of the descriptors of the region's blocks.

The descriptor for each block was obtained as follows. First, each block is normalised to zero mean and unit variance, to reduce the undesired effects of illumination variation. The 2D Discrete Cosine Transform (DCT) [10] is then used as a straightforward dimensionality reduction technique. Specifically, the top ρ low frequency components are retained as the block descriptor, not including the 0-th DCT component (as it has no information due to the normalisation).

Based on preliminary experiments, we used n_R=9, n_p=4 and ρ=15. Note that while the DCT typically decorrelates image data at the block level, there is still correlation among features due to the concatenation of the region descriptors.

As per [16] we report the recognition rates for the 4 illumination sets. The proposed method was compared against Riemannian geodesic distance, Tensor Canonical Correlation Analysis (TCCA) [16] and principal angle [29]. TCCA, as the name implies, is the extension of canonical correlation analysis to multiway data arrays or tensors. Canonical correlation analysis and principal angles are standard methods for measuring the similarity between subspaces [13]. If $\boldsymbol{A} \in \mathbb{R}^{d\times n_1}$ and $\boldsymbol{B} \in \mathbb{R}^{d\times n_2}$ are two linear subspaces in $\mathbb{R}^d$ with minimum rank $r = \min(\text{rank}(\boldsymbol{A}, \boldsymbol{B}))$, then r unique principal angles can be defined between $\boldsymbol{A}$ and $\boldsymbol{B}$ via:

$$\cos(\theta_i) = \max_{\boldsymbol{a}_i\in\boldsymbol{A},\ \boldsymbol{b}_j\in\boldsymbol{B}} \boldsymbol{a}_i^T\boldsymbol{b}_j \tag{13}$$

subject to $\boldsymbol{a}_i^T\boldsymbol{a}_i = \boldsymbol{b}_i^T\boldsymbol{b}_i = 1,\ \boldsymbol{a}_i^T\boldsymbol{a}_j = \boldsymbol{b}_i^T\boldsymbol{b}_j = 0,\ i \neq j$. The principal angle between the two subspaces is $\theta_i \in [0, \pi/2]$, with $i \in \{1, 2, \cdots, r\}$. In line with previous literature [13, 14, 16, 29], we created the subspaces by applying SVD on grey-level images. To compare subspaces, nearest neighbour classification over the first principal angle was employed. The results, presented in Table 1, show that the proposed approach outperforms both the TCCA and principal angle methods by a notable margin.

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

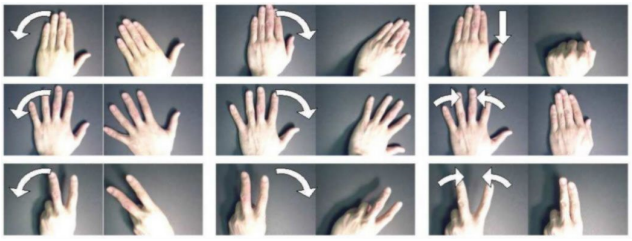

Figure 3. Examples of actions in the Cambridge hand-gesture video dataset [16].

Method	Set1	Set2	Set3	Set4	Overall
Geodesic	66	79	82	85	78.00, σ=8.37
PA [29]	81	74	78	77	77.50, σ=2.89
TCCA [16]	81	81	78	86	81.50, σ=3.32
RLPP	**86**	**86**	**85**	**88**	**86.25**, σ=**1.26**

Table 1. Average correct recognition rate for the hand-gesture recognition task using geodesic distance, principal angle (PA) [29], Tensor Canonical Correlation Analysis (TCCA) [16] and the proposed approach. In the last column, σ represents standard deviation.

5.3. Texture Classification

In this experiment, we performed a classification task using the Brodatz texture dataset [21], which contains 111 texture images of size 640×640. Examples are shown in Fig. 4.

Each image was divided into four equal parts of size 320×320. From each image, we used two parts for training and the remaining two parts for testing. To create a Riemannian manifold, from each 320×320 image, we extracted one hundred rectangular regions of random centre, height and width. We confined the width and height of the regions to be in the range of $[16, 128]$. For every pixel $I(x, y)$ in a region, we then computed a feature vector $F(x,y)=\left[I(x,y), \left|\frac{\partial I}{\partial x}\right|, \left|\frac{\partial I}{\partial y}\right|, \left|\frac{\partial^2 I}{\partial x^2}\right|, \left|\frac{\partial^2 I}{\partial y^2}\right|\right]$. Each region is then described by a 5×5 covariance descriptor of these features.

In the test protocol, for any covariance descriptor we find the nearest neighbour descriptor from the training set and assign the corresponding image class to it. As a result, each 320×320 image is described by one hundred of such labels. The class of each image was obtained using a majority voting rule. Since there are $111 \times 2 \times 100 = 22,200$ points in the training set, generating affinity graphs is computationally intensive. Instead, we randomly select 10 samples from each texture class and train the model using the smaller subset of 10×100 points. Upon deriving the Laplacian space, we projected both the training and testing sets into the new space.

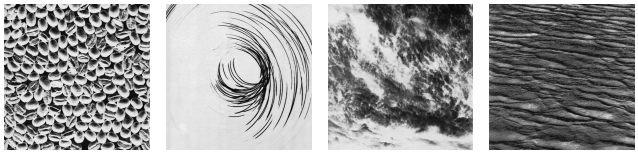

Figure 4. Representative examples from the Brodatz texture dataset [21].

Method	Performance
Maximum response-M8 [9]	94.64%
Leung-Malik [17]	97.32%
Covariance descriptor	95.49%
Geodesic	97.77%
RLPP	99.54%

Table 2. Average correct recognition rate for the texture classification task using maximum response filter bank) [9], Leung-Malik filter bank [17], covariance descriptor, Riemannian geodesic distance and the proposed RLPP approach.

State-of-the-art methods for texture classification utilise the notion of bag of words [22, 31]. More specifically, textons can be considered as visual words derived through clustering a feature space. The feature space is built from the output of a filter bank applied at every pixel, with the methods mainly differing in the employed filter bank. Leung-Malik (LM) [17] and maximum response (MR) [9] filter banks have been shown to be quite successful over the Brodatz dataset [27] and hence are considered here.

The LM filter bank is a combination of 48 anisotropic and isotropic filters and produces a 48 dimensional feature space. The MR filter bank is derived from both rotationally symmetric and oriented filters. To achieve rotational invariance, the responses of the oriented filters are aggregated by a maximum operation. The feature space is 8 dimensional.

Results in Table 2 indicate that the proposed RLPP approach obtains the highest recognition accuracy.

5.4. Person Re-identification

For the person re-identification task, we used the modified ETHZ dataset [23]. The original ETHZ was captured using a moving camera [6], providing a range of variations in appearance of people. The dataset is structured into three sequences. Sequence 1 contains 83 pedestrians (4,857 images), Sequence 2 contains 35 pedestrians (1,936 images), and Sequence 3 contains 28 pedestrians (1,762 images). See Fig. 5 for examples.

We downsampled all images to 64×32 pixels. For each subject we randomly selected 10 images for training and used the rest for testing. Random selection of training and testing data was repeated 20 times to obtain reliable statistics. To describe each image, the covariance descriptor was computed using the following features:

$$F_{x,y}=\left[x, y, R_{x,y}, G_{x,y}, B_{x,y}, R'_{x,y}, G'_{x,y}, B'_{x,y}, R''_{x,y}, G''_{x,y}, B''_{x,y}\right]$$

where x and y represent the position of a pixel, while $R_{x,y}$, $G_{x,y}$ and $B_{x,y}$ represent the corresponding colour information. Furthermore, $C'_{x,y}=\left[\left|\frac{\partial C}{\partial x}\right|, \left|\frac{\partial C}{\partial y}\right|\right]$ and $C''_{x,y}=\left[\left|\frac{\partial^2 C}{\partial x^2}\right|, \left|\frac{\partial^2 C}{\partial y^2}\right|\right]$ represent the gradient and Laplacian for colour C, respectively.

We compared the proposed RLPP method with Partial Least Squares (PLS) [23], Histogram Plus Epitome (HPE) [1], and Symmetry-Driven Accumulation of Local Features (SDALF) [7]. The results are shown in Fig. 6 in terms of recognition rate, by the Cumulative Matching Characteristic (CMC) curve. The CMC curve represents the expectation of finding the correct match in the top n matches. The proposed method obtains superior performance on Sequences 1 and 2, while matching the SDALF method on Sequence 3.

6. Main Findings and Future Directions

Inference problems on Riemannian manifolds are typically tackled by embedding the manifolds into Euclidean spaces. The general practice in this school of thought is to use tangent spaces for embedding. In this paper we proposed a new approach for making inference method on Riemannian manifolds. Specifically, we devised a Riemannian pseudo kernel and employed it for embedding Riemannian manifolds into the familiar RKHS space. To demonstrate the benefits of embedding into RKHS, we recast a locality preserving projection approach from Euclidean spaces to Riemannian manifolds.

When compared to several state-of-the-art methods, experiments on gesture recognition, person re-identification and texture classification indicate that the proposed kernel-based embedding approach leads to considerable improvements in discrimination accuracy.

Future avenues of research include exploring clustering through kernel analysis on Riemannian manifolds. This is particularly useful for creating visual dictionaries over Riemannian manifolds and can open new paths to adapt the ideas of sparse representation [5] to non-Euclidean spaces.

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.

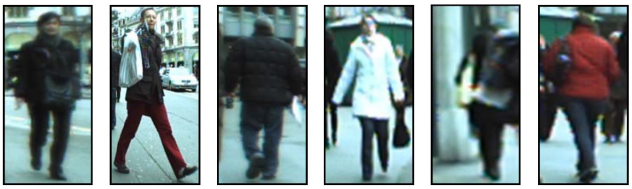

Figure 5. Examples of pedestrians in the ETHZ dataset [6].

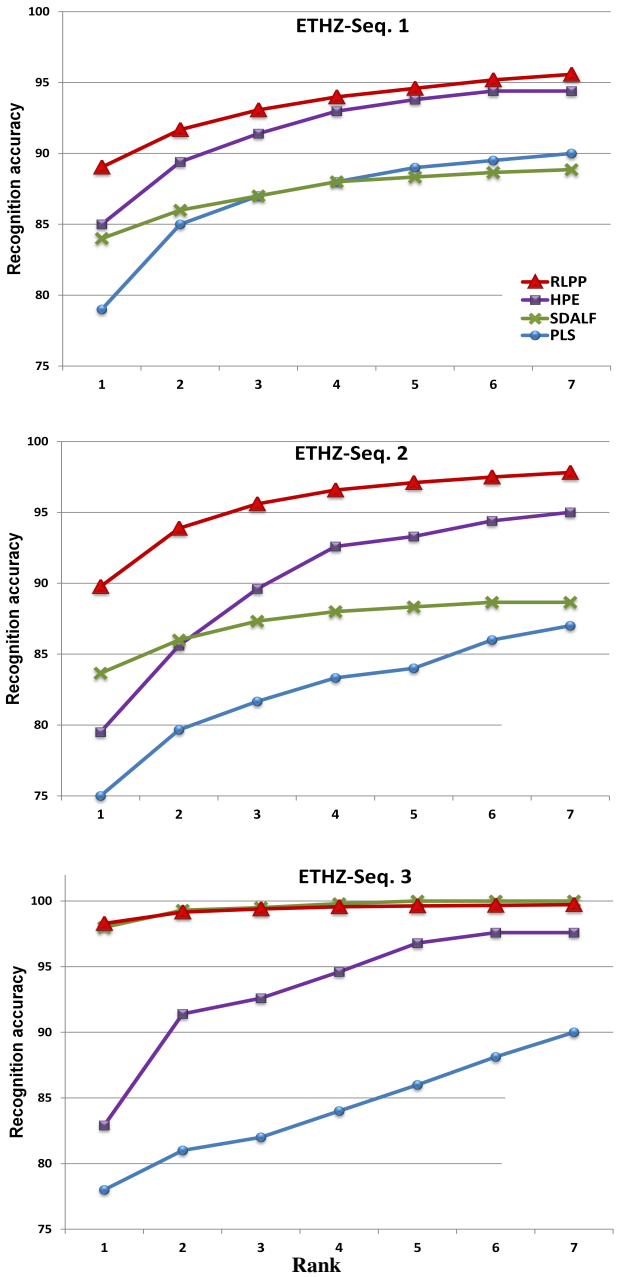

Figure 6. Performance comparison on Sequences 1 through 3 of the ETHZ dataset, in terms of Cumulative Matching Characteristic curves. The proposed RLPP method is compared with Histogram Plus Epitome (HPE) [1], Symmetry-Driven Accumulation of Local Features (SDALF) [7] and Partial Least Squares (PLS) [23].

References

[1] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino. Multiple-shot person re-identification by HPE signature. In *Int. Conf. Pattern Recognition (ICPR)*, pages 1413–1416, 2010.

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

[3] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti. Similarity-based classification: Concepts and algorithms. *Journal of Machine Learning Research*, 10:747–776, 2009.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[5] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[6] A. Ess, B. Leibe, and L. V. Gool. Depth and appearance for mobile scene analysis. *Int. Conf. Computer Vision (ICCV)*, pages 1–8, 2007.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. *IEEE Conf. Computer Vision and Pattern Recognition*, pages 2360–2367, 2010.

[8] J. Friedman, T. Hastie, and R. Tibshirani. Additive Logistic Regression: a Statistical View of Boosting. *The Annals of Statistics*, 28(2):337–407, 2000.

[9] J.-M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic Gauss filtering. *IEEE Trans. Image Processing*, 12(8):938–943, 2003.

[10] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, 2007.

[11] K. Guo, P. Ishwar, and J. Konrad. Action recognition using sparse representation on covariance manifolds of optical flow. In *IEEE Conf. Advanced Video and Signal Based Surveillance (AVSS)*, pages 188–195, 2010.

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

[13] M. Harandi, A. Bigdeli, and B. Lovell. Image-set face recognition based on transductive learning. In *IEEE Int. Conf. Image Processing (ICIP)*, pages 2425–2428, 2010.

[14] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 2705–2712, 2011.

[15] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems 16*. 2004.

[16] T.-K. Kim and R. Cipolla. Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(8):1415–1428, 2009.

[17] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Computer Vision*, 43(1):29–44, 2001.

[18] Y. M. Lui. Advances in matrix manifolds for computer vision. *Image and Vision Computing*, (in press). doi: 10.1016/j.imavis.2011.08.002.

[19] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[20] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on Lie algebra. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 728–735, 2006.

[21] T. Randen and J. H. Husøy. Filtering for texture classification: A comparative study. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.

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

[23] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 322–329, 2009.

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

[25] R. Subbarao and P. Meer. Nonlinear mean shift over Riemannian manifolds. *Int. J. Computer Vision*, 84(1):1–20, 2009.

[26] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(11):2273–2286, 2011.

[27] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *Lecture Notes in Computer Science*, volume 3952, pages 589–600, 2006.

[28] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, 2008.

[29] O. Yamaguchi, K. Fukui, and K.-i. Maeda. Face recognition using temporal image sequence. In *IEEE Conf. 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 Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[31] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Computer Vision*, 73(2):213–238, 2007.