

# Relational Divergence Based Classification on Riemannian Manifolds

Azadeh Alavi, Mehrtash T. Harandi, Conrad Sanderson

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

## Abstract

*A recent trend in computer vision is to represent images through covariance matrices, which can be treated as points on a special class of Riemannian manifolds. A popular way of analysing such manifolds is to embed them in Euclidean spaces, a process which can be interpreted as warping the feature space. Embedding manifolds is not without problems, as the manifold structure may not be accurately preserved. In this paper, we propose a new method for analysing Riemannian manifolds, where embedding into Euclidean spaces is not explicitly required. To this end, we propose to represent Riemannian points through their similarities to a set of reference points on the manifold, with the aid of the recently proposed Stein divergence, which is a symmetrised version of Bregman matrix divergence. Classification problems on manifolds are then effectively converted into the problem of finding appropriate machinery over the space of similarities, which can be tackled by conventional Euclidean learning methods such as linear discriminant analysis. Experiments on face recognition, person re-identification and texture classification show that the proposed method outperforms state-of-the-art approaches, such as Tensor Sparse Coding, Histogram Plus Epitome and the recent Riemannian Locality Preserving Projection.*

## 1. Introduction

Symmetric positive definite matrices (SPD) arise in various problems in machine learning and computer vision. They can be used to describe images and videos [7, 17, 22], as they naturally emerge in the form of covariance matrices and therefore provide compact and informative feature descriptors [4]. In addition to capturing feature correlations compactly, covariance matrices are known to be robust to noise [1]. A key aspect of covariance matrices is their natural geometric property [14], i.e., they form a connected Riemannian manifold. As such, the underlying distance and similarity functions might not be accurately defined in Euclidean spaces [9, 29].

While the theory of learning in Euclidean spaces has been extensively developed, extensions to non-Euclidean spaces like Riemannian manifolds have received relatively little attention. This is mainly due to difficulties of handling the Riemannian structure as compared to straightforward Euclidean geometry. For example, the manifold of

SPD matrices is not closed under normal matrix subtraction. As such, efficiently handling this structure is non-trivial, due largely to two main challenges [27]: (i) defining divergence, distance, or kernel functions on covariances is not easy; (ii) the numerical burden is substantial, even for basic computations such as distances and clustering.

To simplify the handling of Riemannian manifolds, inference is traditionally achieved through first embedding the manifolds in higher dimensional Euclidean spaces. A popular choice for embedding manifolds is through considering tangent spaces [16, 19, 29, 30]. The use of tangent spaces is not free from drawbacks. For instance, only distances between points to the tangent pole are equal to true geodesic distances. This restriction might result in inaccurate modelling, as the structure of the manifolds is only partially taken into account [10]. A recent alternative approach is to embed the manifolds into reproducing kernel Hilbert spaces (RKHS) [24], through dedicated Riemannian kernels [8, 10]. While this approach allows employing the multitude of kernel-based machine learning algorithms, existing Riemannian kernels are either only applicable to subtypes of Riemannian manifolds (e.g., Grassmann manifolds) [8], or are pseudo-kernels [10], meaning they do not satisfy all the conditions of true kernel functions [24].

In this paper, we introduce a new way of analysing Riemannian manifolds where embedding into Euclidean spaces or RKHS is not required. Given a set of training points on a manifold, the idea is to employ the similarity between points as features. More specifically, for each query point to be classified, a similarity to each class is obtained, forming a similarity vector. We obtain each similarity with the aid of the recently proposed Stein divergence [26]. The classification task on manifolds is hence converted into a task in the space of similarity vectors, which can be tackled using learning methods devised for Euclidean spaces. Experiments on several vision tasks (person re-identification, face recognition and texture recognition), show that this new approach outperforms several state-of-the-art methods.

We continue the paper as follows. Section 2 provides a brief overview of Riemannian manifolds and associated divergence measures. We then detail the proposed approach in Section 3, and note its computational complexity in Section 4. 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. Manifold Metrics and Divergences

Without delving too deeply into differential geometry and related topics, Riemannian manifolds are differentiable and smooth manifolds endowed with a Riemannian metric, allowing the notion of lengths and angles to be extended from familiar Euclidean space to curved and non-flat spaces [12]. Points on a Riemannian manifold are connected through smooth curves. The curve with the minimum length and its length are called geodesic and geodesic distance, respectively. Attached to every point of a Riemannian manifold there is a vector space that can be thought of as the set of allowed velocities, constraining the movement of the point on the manifold. This vector space is called the tangent space and enables us to perform differential calculus on the manifold.

The Riemannian structure induced by considering the Affine Invariant Riemannian Metric (AIRM) has been shown to be advantageous for analysing SPD matrices [11, 21]. For  $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{++}^d$ , where  $\mathcal{S}_{++}^d$  is the space of positive definite matrices of size  $d \times d$ , the AIRM is defined as:

$$\delta_R(\mathbf{X}, \mathbf{Y}) := \left\| \log \left( \mathbf{Y}^{-\frac{1}{2}} \mathbf{X} \mathbf{Y}^{-\frac{1}{2}} \right) \right\|_F, \quad (1)$$

where  $\log(\cdot)$  is the principal matrix logarithm [26]. Though the Riemannian structure induced by AIRM is attractive in many aspects, it comes with several difficulties. Firstly, the metric is computationally demanding as it essentially needs eigen-decomposition of  $\mathbf{X}$  and  $\mathbf{Y}$ . More importantly, the resulting structure has negative curvature which prevents us from employing conventional learning algorithms for classification purposes.

In this work, to measure dissimilarity between two SPD matrices  $\mathbf{X}$  and  $\mathbf{Y}$ , we use a version of Bregman matrix divergence known as the Stein divergence [26]. The Bregman matrix divergence for SPD matrices is defined as [13]:

$$D_\phi(\mathbf{X}, \mathbf{Y}) \triangleq \phi(\mathbf{X}) - \phi(\mathbf{Y}) - \langle \nabla_\phi(\mathbf{Y}), \mathbf{X} - \mathbf{Y} \rangle, \quad (2)$$

where  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{X}^T \mathbf{Y})$  and  $\phi : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$  is a real-valued, strictly convex and differentiable function. The Bregman divergence in (2) is asymmetric which is undesirable in many applications. The Jensen-Shannon symmetrisation of Bregman divergence is defined as [13]:

$$D_\phi^{JS}(\mathbf{X}, \mathbf{Y}) \triangleq \frac{1}{2} D_\phi \left( \mathbf{X}, \frac{\mathbf{X} + \mathbf{Y}}{2} \right) + \frac{1}{2} D_\phi \left( \mathbf{Y}, \frac{\mathbf{X} + \mathbf{Y}}{2} \right). \quad (3)$$

If  $\phi$  in (3) is chosen as  $-\log(\det(\mathbf{X}))$ , which is the barrier function of semidefinite cone [26], we arrive at the symmetric Stein divergence, also known as the Jensen Bregman Log-Det divergence [4]:

$$J_\phi(\mathbf{X}, \mathbf{Y}) \triangleq \log \left( \det \left( \frac{\mathbf{X} + \mathbf{Y}}{2} \right) \right) - \frac{1}{2} \log(\det(\mathbf{X} \mathbf{Y})). \quad (4)$$

The symmetric Stein divergence is invariant under congruence transformations and inversion [4]. It is computationally less expensive than AIRM, and is related to AIRM in several aspects, such as the sandwiching property (elucidated below) which establishes a bound between the divergence and AIRM [4].

**Property 1.** Let  $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{++}^d$  and  $\Lambda(\mathbf{X} \mathbf{Y}^{-1})$  be the vector of eigenvalues of  $\mathbf{X} \mathbf{Y}^{-1}$ . The following inequality between the symmetric Stein divergence and Riemannian metric exists [26]:

$$J_\phi(\mathbf{X}, \mathbf{Y}) \leq \frac{1}{8} \delta_R^2(\mathbf{X}, \mathbf{Y}) \leq \frac{1}{4} \delta_T(\mathbf{X}, \mathbf{Y}) (J_\phi(\mathbf{X}, \mathbf{Y}) + d \log d),$$

where  $\delta_T(\mathbf{X}, \mathbf{Y}) = \max_{1 \leq i \leq d} \{ |\log \Lambda(\mathbf{X} \mathbf{Y}^{-1})| \}$  is the Thompson metric [28].

## 3. Relational Divergence Classification

We aim to solve classification tasks originally formulated on Riemannian manifolds by avoiding the traditional approach of using tangent spaces, or embedding into RKHS. More specifically, for each query point (an SPD matrix) to be classified, a similarity to each training class is obtained, forming a similarity vector. We obtain each similarity with the aid of the Stein divergence described in the preceding section. The classification task on manifolds is hence converted into a task in the space of similarity vectors, which can be tackled using learning methods devised for Euclidean spaces, such as linear discriminant analysis, support vector machines or  $k$ -nearest neighbours. In this work we have chosen linear discriminant analysis as a starting point.

Let us define a training set  $\mathbb{C}$  from the underlying Riemannian manifold  $\mathcal{S}_{++}^d$  as:

$$\mathbb{C} = \{(\mathbf{C}_1, y_1), (\mathbf{C}_2, y_2), \dots, (\mathbf{C}_n, y_n)\} \quad (5)$$

where  $y_i \in \{1, 2, \dots, m\}$  is a class label, with  $m$  representing the number of classes. We define the similarity between matrix  $\mathbf{C}_i$  and class  $l$  as:

$$s_{i,l} = \frac{1}{N_l} \sum_{j \neq i} J_\phi(\mathbf{C}_i, \mathbf{C}_j) \delta(y_j - l), \quad (6)$$

where  $\delta(\cdot)$  is the discrete Dirac function and

$$N_l = \begin{cases} n_l - 1 & \text{if } y_i = l \\ n_l & \text{otherwise} \end{cases} \quad (7)$$

with  $n_l$  indicating the number of training matrices in class  $l$ .

Using Eqn. (6) we calculate the similarity between  $\mathbf{C}_i$  and all classes, where  $i \in \{1, 2, \dots, n\}$ . Having the similarity values at our disposal, we represent each Riemannian point  $\mathbf{C}_i$  by a similarity pattern:

$$\mathbf{p}_i = [s_{i,1}, s_{i,2}, \dots, s_{i,m}]^T. \quad (8)$$

The classification task on Riemannian manifolds can now be reinterpreted as a learning task in  $\mathbb{R}^m$ . Given the similarity vectors of training data: (cf. Eqn. 5),

$$\mathbb{P} = \{(\mathbf{p}_1, y_1), (\mathbf{p}_2, y_2), \dots, (\mathbf{p}_n, y_n)\} \quad (9)$$

we are interested in finding a machinery to label a query matrix  $\mathbf{C}_q$ , represented by a similarity vector  $\mathbf{p}_q = [s_{q,1}, s_{q,2}, \dots, s_{q,m}]^T$  which contains the similarity to all  $m$  training classes.

One approach is to use  $\mathbb{P}$  to first find a mapping that minimises the intra-class distances between similarity patterns while simultaneously maximising the inter-class distances. This can be achieved by applying linear discriminant analysis [3], to obtain a mapping  $\mathbf{W}^*$ , as follows:

$$\mathbf{W}^* = \operatorname{argmax}_W \operatorname{trace} \left\{ \left[ \mathbf{W} \mathbf{S}_W \mathbf{W}^T \right]^{-1} \left[ \mathbf{W} \mathbf{S}_B \mathbf{W}^T \right] \right\} \quad (10)$$

where  $\mathbf{S}_B$  and  $\mathbf{S}_W$  are the between class and within class scatter matrices:

$$\mathbf{S}_B = \sum_{i=1}^n (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}) (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})^T \quad (11)$$

$$\mathbf{S}_W = \sum_{l=1}^m \sum_{\substack{i \\ y_i=l}} (\mathbf{p}_i - \boldsymbol{\mu}_l) (\mathbf{p}_i - \boldsymbol{\mu}_l)^T \quad (12)$$

In Eqns. (11) and (12),  $\boldsymbol{\mu}_l$  and  $\bar{\boldsymbol{\mu}}$  are the mean of class  $l$  and overall mean of data in  $\mathbb{P}$ , i.e.,  $\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{y_i=l} \mathbf{p}_i$ , and  $\bar{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n N_i \boldsymbol{\mu}_i$ , where  $N_i$  is the number of points in class  $i$  and  $n$  is the total number of points.

The query similarity vector  $\mathbf{p}_q$  can then be mapped into the new space via:

$$\mathbf{x}_q = \mathbf{W}^{*T} \mathbf{p}_q \quad (13)$$

We can now use a straightforward nearest neighbour classifier [3] to assign a class label to  $\mathbf{x}_q$ . We refer to this approach as Relational Divergence Classification (RDC).

## 4. Computational Complexity

Computational complexity of the training algorithm depends on the complexity of two elements: computing similarity vectors of all training data  $\mathbb{P}$  (Eqn. 5) which requires  $O(n^2 d^3)$  operations, and mapping  $\mathbf{W}^*$  (Eqn. 10) which requires  $O(nm^2 + m^3)$  operations ( $O(nm^2)$  operations to compute the within and between scatter matrices and  $O(m^3)$  to solve Eqn. (10) through eigenanalysis). We note that the training can be done offline. The computational complexity of classifying a query point depends on three factors: calculating its similarity vector which requires  $O(nd^3)$  operations, mapping the resulting similarity vector into  $\mathbb{R}^m$  by  $\mathbf{W}^*$  (Eqn. 13) which demands  $O(m^2)$  operations, and finally applying nearest neighbour classifier [3] in  $\mathbb{R}^m$  which can be done with  $O(mn)$  operations. Hence the complexity of classification for a single query data is equal to  $O(nd^3 + m^2 + mn)$ .

## 5. Experiments and Discussion

We compare the performance of the proposed approach with several state-of-the-art methods on three classification tasks: texture classification, face recognition and person re-identification.

### 5.1. Texture Classification

We use the Brodatz texture dataset [20] and follow the test protocol presented in [25]. Nine test scenarios with various number of classes were generated. The test scenarios include 5-texture ('5c', '5m', '5v', '5v2', '5v3'), 10-texture ('10', '10v') and 16-texture ('16c', '16v') mosaics. See Fig. 1 for examples.

To create a Riemannian manifold, each image was down-sampled to  $256 \times 256$ , followed by splitting into 64 regions of size  $32 \times 32$ . A feature vector for each pixel  $I(x, y)$  is calculated as  $F(x, y) = \left[ I(x, y), \left| \frac{\partial I}{\partial x} \right|, \left| \frac{\partial I}{\partial y} \right|, \left| \frac{\partial^2 I}{\partial x^2} \right|, \left| \frac{\partial^2 I}{\partial y^2} \right| \right]$ . Each region is described by a  $5 \times 5$  covariance matrix formed from these vectors. For each test scenario, 25 covariance matrices per class were randomly selected to construct training data and the rest was used for testing. The random selection was repeated 10 times and the mean results are reported.

Fig. 2 shows the performance of the proposed RDC method against several state of the art methods: log-Euclidean sparse representation (logE-SR) [7, 31], Tensor Sparse Coding (TSC) [25] and Locality Preserving Projection (RLPP) [10]. The results indicate that RDC achieves better performance in 7 out of 9 tests.

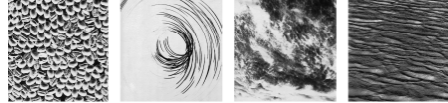


Figure 1. Example images from the Brodatz texture dataset [20].

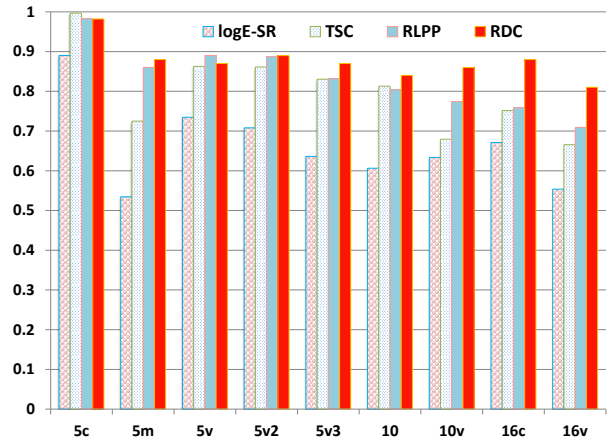


Figure 2. Performance on the Brodatz texture dataset [20] for LogE-SR [7, 31], Tensor Sparse Coding (TSC) [25], Riemannian Locality Preserving Projection (RLPP) [10] and the proposed RDC method.

## 5.2. Face Recognition

For face recognition task, we used the ‘b’ subset of the FERET dataset [18], containing 1400 images of 198 subjects. Each image is closely cropped to include only the face and then downsampled to  $64 \times 64$ . See Fig. 3 for examples.

To evaluate the performance, we created tests with various pose angles. In all tests, training data consisted of the images labelled as ‘ba’, ‘bk’ and ‘bf’ (ie., frontal images with illumination, expression and small pose variations). Images marked as ‘bd’ and ‘bg’ (ie., non-frontal images) were used as separate test sets.

Each face image is represented by a  $43 \times 43$  covariance matrix. To this end, for every pixel  $I(x, y)$  we computed the following feature vector:

$$F_{x,y} = [I(x, y), x, y, |G_{0,0}(x, y)|, \dots, |G_{4,7}(x, y)|] \quad (14)$$

where  $G_{u,v}(x, y)$  is the response of a 2D Gabor wavelet centered at  $x, y$  with orientation  $u$  and scale  $v$  [15]:

$$G_{u,v}(x, y) = \frac{k_v^2}{4\pi^2} \sum_{t,s} e^{-\frac{k_v^2}{8\pi^2}((x-s)^2 + (y-t)^2)} E \quad (15)$$

where  $E = (e^{ik_v((x-t)\cos(\theta_u) + (y-s)\sin(\theta_u))} - e^{-2\pi^2})$ ,  $k_v = \frac{1}{\sqrt{2v-1}}$ , and  $\theta_u = \frac{\pi u}{8}$ .

Table 1 compares the performance of the proposed RDC approach against three state of the art methods: LogE-SR [7], TSC [25] and RLPP [10]. The results show that RDC obtains considerably better accuracy.



**Figure 3.** Examples of closely-cropped faces from the FERET ‘b’ subset [18].

	LogE-SR	TSC	RLPP	RDC
<b>bd</b>	35	36	47	<b>59</b>
<b>bg</b>	47	45	58	<b>71</b>

**Table 1.** Recognition accuracy (in %) for the face recognition task using log-Euclidean sparse representation (logE-SR) [7, 31], Tensor Sparse Coding (TSC) [25], Riemannian Locality Preserving Projection (RLPP) [10] and the proposed RDC method.

## 5.3. Person Re-identification

In this section we test the performance of the proposed RDC method for the person re-identification task on the modified version [23] of the ETHZ dataset [5]. The dataset was captured from a moving camera, with the images of pedestrians containing occlusions and wide variations in appearance. Sequence 1 contains 83 pedestrians (4857 images), while Sequence 2 contains 35 pedestrians (1936 images). See Fig. 4 for examples.

We downsampled all the images to  $64 \times 32$ . For each subject, the training set consisted of 10 randomly selected images, with the rest used for the test set. The random selection of the training and testing data was repeated 20 times. Each image was represented as a covariance matrix formed from feature vectors obtained for each pixel:

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

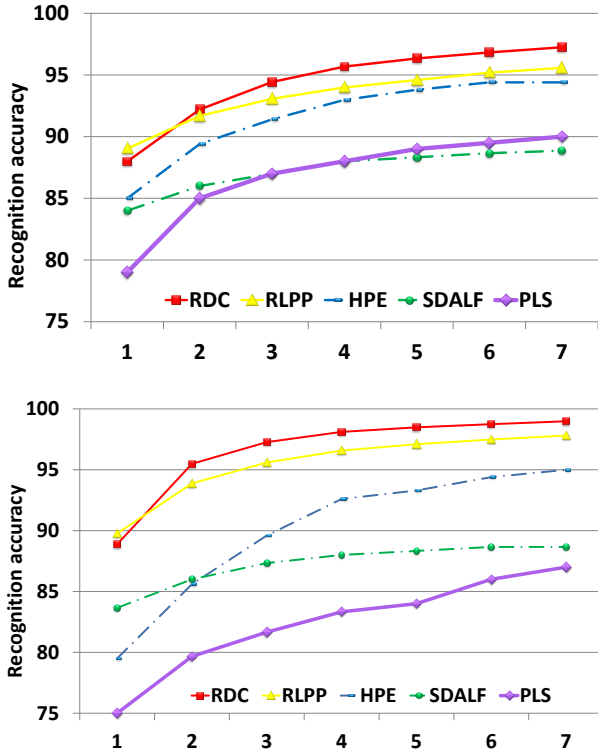
where  $x$  and  $y$  represent the position of a pixel, while  $R_{x,y}$ ,  $G_{x,y}$  and  $B_{x,y}$  represent the corresponding colour information.  $C'_{x,y} = \left[ \left| \frac{\partial C}{\partial x} \right|, \left| \frac{\partial C}{\partial y} \right| \right]$  and  $C''_{x,y} = \left[ \left|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

We compared the proposed RDC with several state-of-the-arts algorithms for person re-identification: Histogram Plus Epitome (HPE) [2], Symmetry-Driven Accumulation of Local Features (SDALF) [6], Partial Least Squares (PLS) [23] and RLPP [10]. The performance of TSC [25] was not evaluated due to the method’s high computational demands: it would take approximately 200 hours on the ETHZ dataset.

The results are shown in Fig. 5, in terms of Cumulative Matching Characteristic (CMC) curves. A CMC curve represents the expectation of finding the correct match in the top  $n$  matches. For  $n > 2$ , the proposed RDC method obtains better performance than the other techniques.



**Figure 4.** Examples of pedestrians in the ETHZ dataset [23].



**Figure 5.** Performance comparison on Sequence 1 (top) and 2 (bottom) of the ETHZ dataset [23], in terms of Cumulative Matching Characteristic curves. The proposed RDC method is compared with Riemannian Locality Preserving Projection (RLPP) [10], Histogram Plus Epitome (HPE) [2], Symmetry-Driven Accumulation of Local Features (SDALF) [6] and Partial Least Squares (PLS) [23].

#### 5.4. Time Comparison

Table 2 compares the total wall-clock time for training and classification required for the proposed RDC approach and the RLPP method [10]. RLPP is a Riemannian manifold based method that employs geodesic distance. The results show that the RDC approach is roughly five times faster than RLPP.

	<b>Brodatz</b>	<b>ETHZ</b>	<b>FERET</b>
<b>RLPP</b>	2.84	4231.09	2343.92
<b>RDC</b>	0.49	921.31	417.68

**Table 2.** Average time (secs) to run the RDC and RLPP algorithms on a 3 GHz machine (Matlab implementation).

## 6. Main Findings and Future Directions

A recent trend in computer vision is to represent images through covariance matrices, which can be treated as points on a special class of Riemannian manifolds. To simplify the handling of such manifolds, inference is traditionally achieved through first embedding the manifolds in higher dimensional Euclidean spaces. While the popular choice for embedding manifolds is through considering tangent spaces, this is not free from drawbacks, as only distances between points to the tangent pole are equal to true geodesic distances. This restriction might result in inaccurate modelling, as the structure of the manifolds is only partially taken into account

In this paper we proposed a new method for analysing Riemannian manifolds, where embedding into Euclidean spaces is not explicitly required. To this end, we proposed to represent Riemannian points through their similarities to a set of reference points on the manifold, with the aid of the recently proposed Stein divergence, which is a symmetrised version of Bregman matrix divergence. Classification problems on manifolds are then effectively converted into the problem of finding appropriate machinery over the space of similarities, which can be tackled by conventional Euclidean learning methods such as linear discriminant analysis.

Experiments on face recognition, person re-identification and texture classification show that the proposed method outperforms state-of-the-art approaches, such as Tensor Sparse Coding, Histogram Plus Epitome, log-Euclidean Sparse Representation, and the recent Riemannian Locality Preserving Projection (RLPP). We have also shown that the proposed method is approximately 5 times faster than RLPP.

In the proposed approach we have currently used all training samples as reference points (see Eqn. 6). While this approach is effective, it might be affected by outliers. Furthermore, for classes with many examples, this approach may not be scalable. As such, future avenues of research include defining alternative methods of creating class representatives, such clustering directly on the manifold.

## 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.

## References

- [1] C. Anoop, M. Vassilios, and P. Nikolaos. Dirichlet process mixture models on symmetric positive definite matrices for appearance clustering in video surveillance applications. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 3417–3424, 2011.
- [2] 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.
- [3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In *Int. Conf. Computer Vision (ICCV)*, pages 2399–2406, 2011.
- [5] A. Ess, B. Leibe, and L. Van Gool. Depth and appearance for mobile scene analysis. In *Int. Conf. Computer Vision (ICCV)*, pages 1–8, 2007.
- [6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 2360–2367, 2010.
- [7] 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.
- [8] J. Hamm and D. D. Lee. Extended Grassmann kernels for subspace-based learning. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 601–608, 2009.
- [9] M. Harandi, C. Sanderson, R. Hartley, and B. Lovell. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *European Conference on Computer Vision (ECCV)*, volume 7573 of *Lecture Notes in Computer Science*, pages 216–229, 2012.
- [10] M. Harandi, C. Sanderson, A. Wiliem, and B. Lovell. Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures. In *IEEE Workshop on the Applications of Computer Vision (WACV)*, pages 433–439, 2012.
- [11] T. Hou and H. Qin. Efficient computation of scale-space features for deformable shape correspondences. *European Conference in Computer Vision (ECCV)*, pages 384–397, 2010.
- [12] I. Kolár, J. Slovák, and P. Michor. *Natural operations in differential geometry*. Springer, 1999.
- [13] B. Kulis, M. Sustik, and I. Dhillon. Low-rank kernel learning with Bregman matrix divergences. *The Journal of Machine Learning Research*, 10:341–376, 2009.
- [14] S. Lang. *Fundamentals of differential geometry*, volume 160. Springer Verlag, 1999.
- [15] T. Lee. Image representation using 2d Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996.
- [16] Y. Lui. Tangent bundles on special manifolds for action recognition. *IEEE Trans. Circuits and Systems for Video Technology*, 22(6):930–942, 2011.
- [17] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
- [18] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [19] 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.
- [20] T. Randen and J. Husoy. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.
- [21] D. Raviv, A. Bronstein, M. Bronstein, R. Kimmel, and N. Sochen. Affine-invariant geodesic geometry of deformable 3d shapes. *Computers & Graphics*, 2011.
- [22] A. Sanin, C. Sanderson, M. Harandi, and B. C. Lovell. Spatio-temporal covariance descriptors for action and gesture recognition. In *IEEE Workshop on the Applications of Computer Vision (WACV)*, 2013.
- [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. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos. Tensor sparse coding for region covariances. In *European Conference on Computer Vision (ECCV)*, volume 6314 of *Lecture Notes in Computer Science*, pages 722–735, 2010.
- [26] S. Sra. Positive definite matrices and the symmetric Stein divergence. *Preprint: [arXiv:1110.1773]*, 2012.
- [27] S. Suvrit and C. Anoop. Generalized dictionary learning for symmetric positive definite matrices with application to nearest neighbor retrieval. *European Conference on Machine Learning (ECML)*, 2011.
- [28] A. C. Thompson. On certain contraction mappings in a partially ordered vector space. *Proceedings of the American Mathematical Society*, 14:438–443, 1963.
- [29] 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.
- [30] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *Pattern Analysis and Machine Intelligence*, 27(12):1896–1909, 2005.
- [31] C. Yuan, W. Hu, X. Li, S. Maybank, and G. Luo. Human action recognition under log-Euclidean Riemannian metric. In *Asian Conference on Computer Vision (ACCV)*, volume 5994 of *Lecture Notes in Computer Science*, pages 343–353. Springer Berlin / Heidelberg, 2010.