Fisher tensors for classifying human epithelial cells

Masoud Faraki [a,b], Mehrtash T. Harandi [c,d,*], Arnold Wiliem [a], Brian C. Lovell [a]

[a] *The University of Queensland, School of ITEE, QLD 4072, Australia*
[b] *Isfahan University of Technology, Department of ECE, Isfahan, Iran*
[c] *NICTA, Locked Bag 8001, Canberra, ACT 2601, Australia*
[d] *Australian National University, Canberra, ACT 0200, Australia*

ARTICLE INFO

ABSTRACT

Analyzing and classifying Human Epithelial type 2 (HEp-2) cells using Indirect Immunofluorescence protocol has been the golden standard for detecting connective tissue diseases such as Rheumatoid Arthritis. However, this suffers from numerous shortcomings such as being subjective as well as time and labor intensive. Recently, several studies explore the advantages of artificial systems to automate the process, not only to reduce the test turn-around time but also to deliver more consistent results. In this paper, we extend the conventional bag of word models from Euclidean space to non-Euclidean Riemannian manifolds and utilize them to classify the HEp-2 cells. The main motivation comes from the observation that HEp-2 cells can be efficiently described by symmetric positive definite matrices which lie on a Riemannian manifold. With this motivation, we first discuss an intrinsic bag of Riemannian words model. We then propose Fisher tensors which can in turn encode additional information about the distribution of the signatures in a bag of word model. Experiments on two challenging HEp-2 images datasets, namely ICPRContest and SNPHEp-2 show that the proposed methods obtain notable improvements in discrimination accuracy, in comparison to baseline and several state-of-the-art methods. The proposed framework, while hand-crafted towards cell classification, is a generic framework for object recognition. This is supported by assessing the performance of our proposal on a challenging texture classification task.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Indirect Immunofluorescence protocol applied on Human Epithelial Cell Type 2 (HEp-2 cells) is considered the standard way of detecting autoimmune diseases such as Systemic Lupus Erythematosus (SLE) and Rheumatoid Arthritis [1,2]. However, this protocol is labor intensive and time consuming as at least two scientists are required in the analytic step [1]. Recently, several studies have proposed cell classification systems to automatically classify HEp-2 cell images, not only to reduce the test turn-around time but also to deliver more consistent results (for example to adjudicate any discrepancy opinion between scientists) [3–12].

In this paper, we utilize COVariance Descriptor (CovD) in Bag of visual Words (BoW) [13] framework for identifying HEp-2 cell images. CovD is a generic and relatively novel image descriptor, introduced by Tuzel et al. [14] and has been successfully employed in a wide variety of visual recognition tasks including pedestrian detection [15], tracking non-rigid moving objects [16], analysing

diffusion tensor images [17], face recognition [18], and action and gesture recognition [19].

Despite their appealing properties, CovDs lie on a special type of Riemannian manifold, the manifold of Symmetric Positive Definite (SPD) matrices or tensors. This in turn makes analyzing and developing inference methods on CovDs quite challenging. In recent years, attempts to generalize machineries originally designed for vector spaces to their Riemannian counterparts have received a surge of attention [15,17,19–25]. Following such trends, in this work we discuss how BoW models can be extended to their Riemannian versions.

To this end, we first elaborate on an intrinsic Bag of Riemannian Words (BoRW) model (see Section 4 for the definition of intrinsic methods). The intrinsic BoRW model takes into account the true geometry of tensors in obtaining its codebook and signatures.[1] We then introduce our main player in this work, Fisher Tensors (FT), which is the Riemannian version of Fisher vectors [26] and can encode additional information about the distribution of the signatures (as compared to BoRW). This is motivated by the recent success of Fisher vectors [26], especially on large-scale object recognition tasks, over conventional BoW models [26–28]. In FT,

---

* Corresponding author at: Australian National University, Canberra, ACT 0200, Australia. Tel.: +61 262676252.

E-mail addresses: m.faraki@ec.iut.ac.ir (M. Faraki),
mehrtash.harandi@nicta.com.au (M.T. Harandi), a.wiliem@uq.edu.au (A. Wiliem),
lovell@itee.uq.edu.au (B.C. Lovell).

[1] We use signature and histogram in BoW interchangeably in this paper.

local descriptors encoded by CovD are soft-assigned to components of a Gaussian Mixture Model (GMM) and the image is represented using the gradient of the log-likelihood of the local CovDs with respect to the GMM. To achieve this, a probabilistic model (in this work GMM) should be defined over Riemannian manifolds. Our idea here is to utilize a diffeomorphism and define the probabilistic model by embedding the Riemannian manifold into a vector space. This leads to an elegant model with a significantly smaller codebook as compared to conventional BoW models. Rigorous tests on two challenging HEp-2 cell datasets, namely ICPR2012Contest [29] and SNPHEp-2 [12] show the efficiency and power of the proposed FT for classifying HEp-2 cells and texture images.

Contributions: In summary, there are five novelties in this work.

1. We propose the use of CovDs for analyzing indirect immuno-fluorescence images.
2. We elaborate on an intrinsic BoRW model.
3. We propose Fisher tensors, a mechanism for utilizing Fisher vectors on Riemannian manifolds.
4. We compare and contrast our proposed algorithms against several state-of-the-art methods (e.g. the methods proposed by Cordelli and Soda [9], Strandmark et al. [30], and Wiliem et al. [12]) on HEp-2 Cells classification task. Our empirical results show that the proposed approach obtains superior performance on two challenging and recently released HEp-2 Cells datasets, ICPR2012Contest and SNPHEp-2.
5. Finally, though in this work, FTs are hand-crafted towards classifying the HEp-2 cells, the proposed framework is a generic object classification. We therefore, consider the task of textures classification and show the superiority of FTs over the-state-of-the-art methods on KTH-TIPS [31] texture dataset.

We continue this paper as follows. Section 2 introduces recent advances in classifying HEp-2 cells and briefly reviews BoW models and Fisher vectors. Section 3 is dedicated to Riemannian geometry and serves as a grounding for follow-up sections. Section 4 discusses the intrinsic bag of Riemannian words model. Section 5 introduces the Fisher tensors, the Riemannian version of Fisher vectors. In Section 6 we compare the performance of the proposed method with previous approaches on several datasets. The main findings and possible future directions are summarized in Section 8.

2. Related work

In this section we discuss some of the most relevant work on cell classification systems for HEp-2 cell identification, bag of visual words, and Fisher vectors.

Systems for HEp-2 cell identification: Generally speaking, HEp-2 cells can be recognized either globally or locally. In the global methods, cell images are described by holistic features such as grey level statistics and textural features [7–11,30], whilst in the latter, statistics of local features (extracted from local patches) are used for recognition [12,32]. In addition, the holistic category can be further divided into two sub-categories, namely semi-automatic systems based on hand-picked features, and fully automatic systems that exploit feature selection methods. In the semi-automatic systems, expert knowledge and human interactions are required for selecting discriminative features (i.e., manual feature selection) [7–11,30]. One example is the work of Elbischger et al. where image statistics, cell shape and textural features, in conjunction with Nearest Neighbour (NN) classifier, are used to identify HEp-2 cells [7]. Recently, Strand-mark et al. proposed ensemble learning on hand-picked features for recognising HEp-2 cells [30]. Methods classified within this category are prone to inter-laboratory settings (i.e., each laboratory may have

different assays, sample preparation protocols, and acquisition hardware) [12].

To address this problem, automatic feature selection has been proposed [5,4,6]. The majority of algorithms in this category first extract a large set of features and then narrow them down to the most informative ones using feature selection methods. For example, Hiemann et al. extracted 1400 holistic features including textures and morphological features and designed an inference machinery by applying various feature selection methods [5]. Soda and Iannello proposed a multiple expert system (MES) where each system is designed and tuned to classify a specific pattern [4]. The confidence value of each sub-system is then combined to produce the final result, which can be understood as an indirect way of selecting the best features for each class. Compared to hand-picked feature systems, approaches based on automatic feature selection might be more robust to inter-laboratory settings since the system can always be retrained to a new setting by redoing the feature selection process. We refer interested readers to [33] which discusses several automatic HEp-2 cell classification systems.

Bag of visual words: Different from holistic descriptors which come with rigid spatial assumptions on images, it is possible to represent an image by a histogram of local features (i.e., BoW approach) [13]. The core idea of BoW methods is to learn a dictionary of visual words, or codebook, where visual words capture features or image patches that are more likely to occur or discriminative for the task in hand. After learning the dictionary, an image is represented by a histogram where each bin of the histogram shows the frequency of seeing the corresponding visual word in the image. Generally speaking, unlike holistic approaches, there is no need to hand pick or select features in BoW methods since robust and discriminative features can be learnt during the process [34–36]. Moreover, BoW methods have the reputation of being robust to image deformations such as image rotation and affine transformation compared to traditional holistic approaches [37,38].

Fisher vectors: Despite their tremendous simplicity and overwhelming power, conventional BoW models suffer from lack of order statistics in the signatures and a lossy quantization process. More specifically, in BoW model, each local descriptor is assigned to the nearest codeword. This can be understood as encoding the zero-order statistics from a probabilistic modeling perspective and has been studied in [26–28]. Thus, a curious mind might seek methods that bring higher-order statistics into the consideration. This school of thought leads us to the notion of Fisher Vector (FV) [26–28] which usually outperforms conventional BoW models. We note that while it is well established that large codebooks can improve the result in BoW models, assigning local descriptors to their nearest codewords is computationally expensive. Last but not least, in BoW model the descriptor quantization is a lossy process as studied for example in the work of Boiman et al. [39]. In FV, quantization will be replaced by computing the deviation of the vectors with respect to a generative probability model.

3. Background theory

In this section, we review Riemannian geometry of SPD matrices which in turn provides the grounding for techniques described in the following sections. More involved treatments on manifolds and related topics can be found in [40–42]. Throughout this paper we will use the following notation:

- $GL(d)$ is the general linear group, the group of real invertible $d \times d$ matrices.
- $Sym(d)$ is the space of real $d \times d$ symmetric matrices.
- $\mathcal{S}_{++}^d$ is the space of real $d \times d$ SPD matrices.

- $Diag(\lambda_1, \lambda_2, ..., \lambda_d)$ is a diagonal matrix formed from real values $\lambda_1, \lambda_2, ..., \lambda_d$ on diagonal elements.
- The principal matrix logarithm $\log(\cdot)$ is defined as $\log(X) = \sum_{r=1}^{\infty}((-1)^{r-1})/r(X-\mathbb{I})^r = U\,\mathrm{Diag}(\ln(\lambda_i))U^T$ with
  $$X = U\,\mathrm{Diag}(\lambda_i)U^T.$$
- The principal matrix exponential $\exp(\cdot)$ is defined as $\exp(X) = \sum_{r=0}^{\infty}\frac{1}{r!}X^r = U\,\mathrm{Diag}(\exp(\lambda_i))U^T$ with
  $$X = U\,\mathrm{Diag}(\lambda_i)U^T.$$

Before delving more into Riemannian geometry, we need to formally define CovDs. Let $f(x, y)$ be a $W \times H$ greyscale image.[2] Let $\mathbb{O} = \{\boldsymbol{o}_i\}_{i=1}^n, \boldsymbol{o}_i \in \mathbb{R}^d$ be a set of observations extracted from $f(x, y)$. For example one might extract a d dimensional feature vector (such as pixel value, gradients along horizontal and vertical directions, and filter responses) at each pixel, resulting in $W \times H$ observations. Then, the image $f(x, y)$ can be represented by a $d \times d$ covariance matrix of the observations as

$$C_f = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{o}_i - \mu)(\boldsymbol{o}_i - \mu)^T,$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{o}_i. \tag{1}$$

The entries on the diagonal of matrix C_f are the variances of each feature and the nondiagonal entries are their pairwise correlations. There are several reasons as to why CovDs are attractive for representing images. First, CovD provides a natural and seamless way for fusing various features. Second, it can reduce the impact of noisy samples in a region through the averaging operation inherently involved in its computation. Third, a $d \times d$ CovD is usually low-dimensional and independent of the size of the region (it has only $\frac{1}{2}d(d+1)$ unique values), therefore it can be easily utilized to compare regions with various sizes. Fourth, CovDs can be efficiently computed using integral images (see [15] for 2D integral images and [19] for 3D ones). Finally, affine invariant metrics exist to compare CovD which is of special significance in computer vision tasks [17,43].

3.1. Riemannian geometry of SPD matrices

The space of real $d \times d$ SPD matrices $\mathcal{S}_{++}^d$, forms a Lie Group, an algebraic group with a manifold structure. It is natural to use the language of Riemannian manifolds, geodesics and all the related concepts of differential geometry when discussing $\mathcal{S}_{++}^d$.

A manifold is a locally Euclidean Hausdorff space whose topology has a countable base. Locally Euclidean just means that each point has some neighborhood that is homeomorphic to an open ball in $\mathbb{R}^d$ for some d. Moreover, being a Hausdorff space means that distinct points have disjoint neighbourhoods. This property is useful for establishing the notion of a differential manifold, as it guarantees that convergent sequences have a single limit point.

A natural way to measure nearness on a manifold is by considering the geodesic distance between points. The geodesic distance between two points on the manifold is defined as the length of the shortest curve connecting the two points. Such shortest curves are known as geodesics and are analogous to straight lines in Euclidean space $\mathbb{R}^d$. The tangent space at a point $\boldsymbol{P}$ on the manifold, $T_{\boldsymbol{P}}\mathcal{M}$, is a vector space that consists of the tangent vectors of all possible curves passing through $\boldsymbol{P}$. Two operators, namely the exponential map $\exp_{\boldsymbol{P}}(\cdot): T_{\boldsymbol{P}}\mathcal{M} \to \mathcal{M}$ and the logarithm map

$\log_{\boldsymbol{P}}(\cdot) := \exp_{\boldsymbol{P}}^{-1}(\cdot): \mathcal{M} \to T_{\boldsymbol{P}}\mathcal{M}$, are defined over differentiable manifolds to switch between the manifold and tangent space at $\boldsymbol{P}$ (see Fig. 1 for a conceptual illustration). The exponential operator maps a tangent vector Δ to a point $\boldsymbol{X}$ on the manifold. The property of the exponential map ensures that the length of Δ becomes equal to the geodesic distance between $\boldsymbol{X}$ and $\boldsymbol{P}$. The logarithm map is the inverse of the exponential map and maps a point on the manifold to the tangent space $T_{\boldsymbol{P}}$. The exponential and logarithm maps vary as point $\boldsymbol{P}$ moves along the manifold.

The Affine Invariant Riemannian Metric (AIRM) [17] defined as

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle_{\boldsymbol{P}} := \langle \boldsymbol{P}^{-1/2}\boldsymbol{v}\boldsymbol{P}^{-1/2}, \boldsymbol{P}^{-1/2}\boldsymbol{w}\boldsymbol{P}^{-1/2} \rangle = \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{v}\boldsymbol{P}^{-1}\boldsymbol{w}), \tag{2}$$

for $\boldsymbol{P} \in \mathcal{S}_{++}^d$ and $\boldsymbol{v}, \boldsymbol{w} \in T_{\boldsymbol{P}}\mathcal{M}$, induces the following geodesic distance between points $\boldsymbol{X}$ and $\boldsymbol{Y}$:

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

Amongst several interesting properties (see for example [17,41]), the invariance to affine transforms, i.e., $\forall \boldsymbol{A} \in GL(d), \ \delta_R^2(\boldsymbol{X}, \boldsymbol{Y}) = \delta_R^2(\boldsymbol{A}\boldsymbol{X}\boldsymbol{A}^T, \boldsymbol{A}\boldsymbol{Y}\boldsymbol{A}^T)$, is especially attractive to computer vision community. For the AIRM, the exponential and logarithms maps are given by

$$\exp_{\boldsymbol{P}}(\boldsymbol{X}) := \boldsymbol{P}^{1/2}\exp(\boldsymbol{P}^{-1/2}\boldsymbol{X}\boldsymbol{P}^{-1/2})\boldsymbol{P}^{1/2}. \tag{4}$$

$$\log_{\boldsymbol{P}}(\boldsymbol{X}) = \boldsymbol{P}^{1/2}\log(\boldsymbol{P}^{-1/2}\boldsymbol{X}\boldsymbol{P}^{-1/2})\boldsymbol{P}^{1/2}. \tag{5}$$

4. Bag of Riemannian words

The BoW method is probably the most popular way to categorize images these days. In a nutshell, the BoW representation of an image is obtained by first extracting local descriptors from the image followed by assigning each descriptor to its closest visual word in a "visual vocabulary" or codebook. The codebook is obtained offline by clustering a large set of descriptors with (usually) k-means algorithm.

In this section we discuss how a conventional BoW model can be extended to incorporate the Riemannian structure of CovD descriptors. While our treatment here is directed towards the geometry of SPD matrices, concepts from this section can be easily generalized to any type of Riemannian manifolds. To devise a BoRW model, we should address two sub-problems:

1. Given a set of points on a SPD manifold (each point on the manifold corresponds to a CovD), how can a Riemannian codebook be trained?
2. Given a Riemannian dictionary, how can a signature be obtained for a query image?
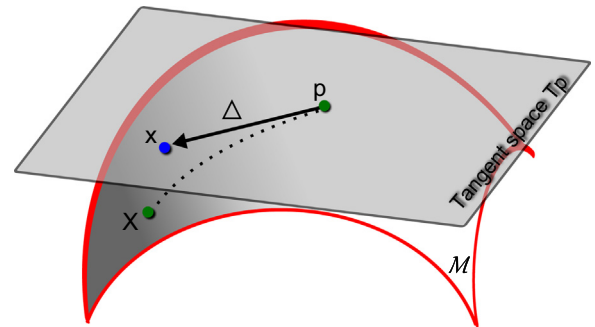
Fig. 1. Illustration of the tangent space $T_{\boldsymbol{P}}$ at point $\boldsymbol{P}$ on a Riemannian manifold $\mathcal{M}$. The tangent vector Δ can be obtained through the logarithm mapping, i.e. $\Delta = \log_{\boldsymbol{P}}(\boldsymbol{X})$. Every tangent vector in $T_{\boldsymbol{P}}$ can be mapped back to the manifold through the exponential map, i.e., $\exp_{\boldsymbol{P}}(\Delta) = \boldsymbol{X}$. The dotted line shows the geodesic starting at $\boldsymbol{P}$ and ending at $\boldsymbol{X}$.

---

[2] Extension of CovD to color images, hyperspectral images or even videos is straightforward.

The conceptual diagram of BoRW signature computation is illustrated in Fig. 2.

4.1. Riemannian codebook

We note that SPD matrices form a closed set under normal matrix addition, i.e., adding two SPD matrices results in another SPD matrix. Therefore, in the simplest and most straightforward form, one could neglect the Riemannian geometry of SPD matrices and vectorize SPD matrices by extracting upper-triangular (or equivalently lower-triangular) part of them. Having data vectorized, then a codebook can be generated by applying k-means algorithm over vectorized data. As a result, for each cluster, the cluster center is given by the arithmetic mean of nearest training samples to that cluster. However, the arithmetic mean is not adequate in many situations, for two main reasons. First, symmetric matrices with negative or zero eigenvalues are at a finite distance from any SPD matrix in this framework. In many problems like diffusion tensor MRI, this is not physically acceptable [17,44]. Second, in our application an SPD matrix corresponds to a covariance matrix. The value of the determinant of a covariance matrix is a measure of the dispersion of the associated multivariate Gaussian. As shown for example by Pennec the arithmetic mean of SPD matrices often leads to a swelling effect [17], that is the determinant of arithmetic mean could become greater than the samples' determinants which is very undesirable [44]. As such, we seek possibilities to take into account the geometry of SPD matrices in designing a codebook.

Let $\mathbb{X} = \{X_i\}_{i=1}^N$ be a set of training samples from the underlying $\mathcal{S}_{++}^d$ manifold. A codebook $\mathbb{D} = \{D_i\}_{i=1}^k$, $D_i \in \mathcal{S}_{++}^d$ on $\mathcal{S}_{++}^d$ can be obtained through intrinsic k-means algorithm [17]. The core ingredient of intrinsic k-means algorithm is the Karcher mean [17]. Given an abstract manifold $\mathcal{M}$ with associated geodesic distance $d_g : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$, the classical generalization of the Euclidean mean for a set of points $\{X_i\}_{i=1}^m$, $X_i \in \mathcal{M}$ is given by the Karcher mean (also referred as Fráchet or Riemannian mean) as the point minimizing the following metric dispersion:

$$X^* = \arg \min_X \sum_{i=1}^m d_g^2(X_i, X). \tag{6}$$

Since the manifold of SPD matrices has a non-positive curvature, (6) has just one and only one optimum [41]. Moreover, at optimum, the gradient is zero and therefore an intrinsic gradient descent algorithm can be utilized to obtain the Karcher mean [17]. The details of computing the Karcher mean over SPD manifolds are given in Algorithm 1.

Having an intrinsic method for computing mean at our disposal, we seek to estimate k clusters $C_1, C_2, ..., C_k$ with centers $\{D_i\}_{i=1}^k$ such that the sum of square geodesic-distances, i.e., $\sum_{i=1}^k \sum_{X_j \in C_i} d_g^2(X_j, D_i)$, over all clusters is minimized. Similar to standard k-means, this can be solved using an EM-based approach. The algorithm starts by selecting k points from $\mathbb{X}$ randomly as the cluster centers. In the E-step, we assign each of the points of the dataset to the nearest cluster center. Then in the M-step, the cluster centers are recomputed using the Karcher mean. The procedure is summarized in Algorithm 2.

Algorithm 1. Karcher mean algorithm over $\mathcal{S}_{++}^d$.

Input:
- A set of points $\{X_i\}_{i=1}^m$ on the underlying $\mathcal{S}_{++}^d$ manifold,
- *maxIter*, maximum number of iterations

Output:
- The sample Karcher mean μ

1: Let $\mu^{(0)}$ be an initial estimate of the Karcher mean, for example by selecting a sample from $\{X_i\}_{i=1}^m$ randomly.
2: **while** $t < maxIter$ **do**
3: 　　For each point X_i, compute the tangent vector v_i on the current estimate of the Karcher mean, i.e., $v_i = \log_{\mu^{(t)}}(X_i)$.
4: 　　Compute the average tangent vector $\bar{v} = \frac{1}{k} \sum_{i=1}^m v_i$.
5: 　　If $\|\bar{v}\|_2$ is small, then $\mu = \exp_{\mu^{(t)}}(\bar{v})$ and break the while. Else, compute the new estimate of Karcher mean $\mu^{(t+1)}$ by moving along the average tangent direction, i.e., $\mu^{(t+1)} = \exp_{\mu^{(t)}}(\varepsilon\bar{v})$, where $\varepsilon > 0$ is small step size.
6: 　　$t \leftarrow t+1$
7: **end while**

Algorithm 2. Intrinsic k-means algorithm over $\mathcal{S}_{++}^d$ for learning the visual dictionary.

Input:
- training set $\mathbb{X} = \{X_i\}_{i=1}^N$ from the underlying $\mathcal{S}_{++}^d$ manifold,
- *nIter*, the number of iterations

Output:
- Visual dictionary $\mathbb{D} = \{D_i\}_{i=1}^k$, $D_i \in \mathcal{S}_{++}^d$

1: Initialize the dictionary $\mathbb{D} = \{D_i\}_{i=1}^k$ by selecting N samples from $\mathbb{X}$ randomly
2: **for** $t = 1 \rightarrow nIter$ **do**
3: 　　Assign each point X_i to its nearest cluster in $\mathbb{D}$ by computing $\delta_R(X_i, D_j) = \|\log(X_i^{-1/2} D_j X_i^{-1/2})\|_F$, $1 \leq i \leq N, 1 \leq j \leq k$
4: 　　Recompute cluster centres $\{D_i\}_{i=1}^k$ by Karcher mean algorithm
5: **end for**

4.2. Signatures

In its most straightforward and simplest form, for a set of local CovD, $\mathbb{Q} = \{Q_i\}_{i=1}^p$, extracted from a query image, a signature is obtained by hard-assignment. This demands $p \times k$ comparisons which could be prohibitive for large dictionaries given the high computational cost of AIRM. Recently, the symmetric Stein divergence is proposed as a metric on $\mathcal{S}_{++}^d$ [43]. This metric is defined as

$$S(X, Y) \triangleq \ln \det\left(\frac{X+Y}{2}\right) - \frac{1}{2} \ln \det(XY), \quad \text{for } X, Y > 0, \tag{7}$$

where det($\cdot$) denotes the determinant. Similar to AIRM, the symmetric Stein divergence is invariant to affine transform [43]. Other important properties of symmetric Stein divergence are summarized below.
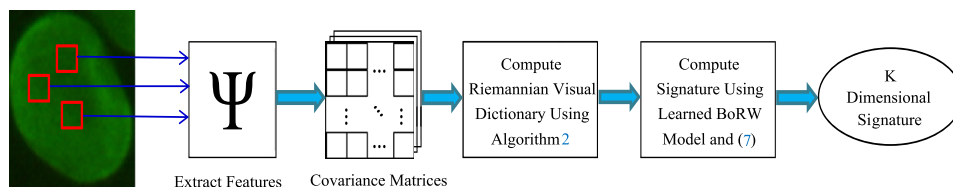

Fig. 2. Block diagram showing computations of BoRW signature.

Property 1. *Let $X, Y \in \mathcal{S}_{++}^d$, and $\delta_T(X, Y) = \max_{1 \leq i \leq d} \{|\log \Lambda(XY^{-1})|\}$ be the Thompson metric [45] with $\Lambda(XY^{-1})$ representing the vector of eigenvalues of XY^{-1}. The following sandwiching inequality between the symmetric Stein divergence and Riemannian metric exists [43]:*

$$S(X, Y) \leq \frac{1}{8} \delta_R^2(X, Y) \leq \frac{1}{4} \delta_T(X, Y)(S(X, Y) + d \log d). \quad (8)$$

Property 2. *The curve $\gamma(p) \triangleq X^{1/2}(X^{-1/2}YX^{-1/2})^p X^{1/2}$ parameterizes the unique geodesic between the SPD matrices X and Y. On this curve the Riemannian geodesic distance satisfies $\delta_R(X, \gamma(p)) = p\delta_R(X, Y); p \in [0, 1]$ [41]. The symmetric Stein divergence satisfies a similar but slightly weaker result, $S(X, \gamma(p)) \leq pS(X, Y)$.*

The first property establishes a bound between the AIRM and Stein divergence, providing motivation for addressing Riemannian problems via the divergence. The second property reinforces the motivation by explaining that the behavior of Stein divergences along geodesic curves is similar to true AIRM metric. In this work, we use the Stein metric instead of AIRM for computing the signatures for BoRW. The computational burden of AIRM as compared to Stein metric has been recently studied by Cherian et al. [46].

5. The Fisher tensors

In the previous section, we introduced the BoRW model. In this section we propose an extension of BoRW which takes into account higher statistics in deriving the image signatures. We start this section by providing a short description of Fisher vectors. We then elaborate how the Fisher vector framework can be extended on the Riemannian manifold of SPD matrices.

Fisher vectors: Given a generative model with probability distribution $p(x|\lambda)$, $x \in \mathbb{R}^d$, Jaakkola et al. introduced the Fisher score $G_\lambda(x) = \nabla_\lambda \log p(x|\lambda)$ as a means to design a Mercer kernel on the generative model [47]. The underlying idea can be explained by considering the class of probability models $p(x|\lambda)$ over λ which defines a Riemannian manifold M_λ. A local metric on M_λ is defined by Fisher information matrix $I(x) = \mathbb{E}\{G_\lambda(x)G_\lambda(x)^T\}$ where $\mathbb{E}$ denotes expectation. Thus, the Fisher score $G_\lambda(x)$ maps an example x from M_λ into a feature vector that is a point in the tangent space of M_λ and basically encodes the direction of the steepest ascent in $\log p(x|\lambda)$ along the manifold M_λ. Later, Perronnin and Dance showed that a powerful image descriptor can be obtained by concatenating Fisher scores into a vector and called it Fisher vectors [26]. We note that the conventional BoW model can be interpreted as applying the Fisher kernel framework [47] to a simple i.i.d. multinomial model over visual word indices [48]. Dwelling more on Fisher scores and related topics is beyond the scope of this paper. Interested readers are referred to [26,28,47,49] and references therein for a detailed treatment.

Fisher tensors: Similar to its Euclidean counterpart, for Fisher tensors we choose $p(X|\lambda)$ to be a GMM [50]. However, in our case, $X \in \mathcal{S}_{++}^d$ and hence we need to derive a generative model on $\mathcal{S}_{++}^d$. A handful of studies discuss how normal distributions and consequently mixture models can be extended to their Riemannian versions intrinsically [17,51]. Our idea here is to simplify the problem by embedding the manifold into a vector space. For this purpose, we make use of a mapping from $\mathcal{S}_{++}^d$ into the space of symmetric matrices by the principal matrix logarithm. The motivation comes from the fact that unlike the general case of invertible square matrices, there always exists a unique, real and symmetric logarithm for any SPD matrix, which can be obtained by principal logarithm. Moreover, $\log(\cdot)$ on $\mathcal{S}_{++}^d$ is diffeomorphism (a one-to-one, continuous, differentiable mapping with a continuous, differentiable inverse). Formally,

Theorem 1. $\log(\cdot): \mathcal{S}_{++}^d \to Sym(d)$ *is C^∞ and therefore both $\log(\cdot)$ and its inverse $\exp(\cdot)$ are smooth, i.e., they are diffeomorphisms.*

Proof. We refer the reader to [44] for the proof of this theorem. □

Embedding into the space of symmetric matrices through principal logarithm can be also understood as embedding $\mathcal{S}_{++}^d$ into its tangent space at identity matrix. Since symmetric matrices (and on a similar note tangent spaces) form a vector space, then we could seamlessly employ Euclidean tools (like Expectation Maximization [50] algorithm to obtain a GMM) to tackle the problem in hand. Other properties of the induced space or the log-Euclidean space (as it is usually referred by in the literature) are studied amongst the others in [44]. Recent applications of the log-Euclidean space include but are not limited to foreground segmentation [52], object tracking [53] and action recognition [54]. We note that our idea here can be labeled as an extrinsic approach, i.e., it depends on the embedding Euclidean space. On the contrary, geodesic distance, the Karcher mean, and intrinsic k-means algorithm depend only on the manifold and therefore they are intrinsic.

Given an SPD matrix A, its log-Euclidean vector representation, a, is unique and defined as $a = Vec(\log(A))$ where the $Vec(B)$, $B \in Sym(d)$ is

$$Vec(B) = [b_{1,1}, \sqrt{2}b_{1,2}, \sqrt{2}b_{1,3}, \cdots \sqrt{2}b_{1,d}, b_{2,2}, \sqrt{2}b_{2,3}, \cdots b_{d,d}]^T. \quad (9)$$

Let $p(x|\lambda)$ be a probabilistic distribution modeled by a GMM with K Gaussians, i.e., $p(x|\lambda) = \sum_{i=1}^K \omega_i g(x|\mu_i, \Sigma_i)$, with ω_i and $g(x|\mu_i, \Sigma_i)$ being the mixture weights and the component Gaussian densities, respectively. Each component density is a $\frac{1}{2}d(d+1)-$variate Gaussian function of the form:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{(d(d+1)/2)}|\Sigma_i|^{1/2}} \exp\left\{\frac{-1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right\}. \quad (10)$$

Given a set of local CovD, $\mathbb{Q} = \{Q_i\}_{i=1}^p$, extracted from a query image, we assume samples are i.i.d. and generated from the probability distribution modelled by $p(x|\lambda)$. This enables us to describe the set $\mathbb{Q} = \{Q_i\}_{i=1}^p$ by a Fisher tensor as the concatenation of K Fisher scores (gradient vectors w.r.t. probability model), $\{\mathcal{G}_i^T\}_{i=1}^K$. That is $FT(\mathbb{Q}) = [\mathcal{G}_1^T, \mathcal{G}_2^T, ..., \mathcal{G}_K^T]^T$, with $q_i = Vec(\log(Q_i))$ and

$$\mathcal{G}_i = \frac{1}{p}\sum_{t=1}^p \nabla_\lambda \log g(q_t|\mu_i, \Sigma_i). \quad (11)$$

With fixed and diagonal covariance matrices, i.e., $\Sigma_i = \sigma_i^2 \mathbb{I}$, let $\gamma_i(q_t)$ be the soft-assignment of log-Euclidean vector q_t to the i-th Gaussian components, i.e.:

$$\gamma_i(q_t) = \frac{\omega_i g(q_t|\mu_i, \Sigma_i)}{\sum_{j=1}^K \omega_j g(q_t|\mu_j, \Sigma_j)}. \quad (12)$$

Then the *ith* Fisher score becomes

$$\mathcal{G}_i = \frac{1}{p\sqrt{\omega_i}}\sum_{t=1}^p \gamma_i(q_t)\sigma_i^{-1}(q_t - \mu_i). \quad (13)$$

We note that in (13), the gradient is taken with regards to μ_i which in turns adds first order statistics into the FT. It is also possible to compute the gradient with respect to covariance matrix Σ_i and hence add second order statistics to the model. However, in our experiments we did not gain significant improvement by exploiting second order statistics. As such, for clarity purposes we confined ourselves to explain the model with first order statistics only. Algorithm 3 and Fig. 3 assembles all the above details into pseudo-code for FT.
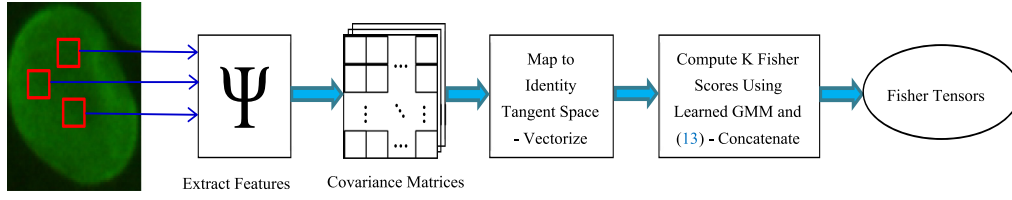
Fig. 3. Block diagram showing Fisher Tensor computation.

Algorithm 3. Fisher Tensor computation algorithm over $\mathcal{S}_{++}^d$.

Input:
- $\mathbb{Q} = \{\mathbf{Q}_t\}_{t=1}^p$, CovDs extracted from a query image,
- $\lambda = \{\lambda_i\}_{i=1}^K$, $\lambda_i = \{\omega_i, \mu_i, \Sigma_i\}$, GMM components

Output:
- $FT(\mathbb{Q})$, Fisher Tensor of $\mathbb{Q}$

1: Compute log-Euclidean representation of $\mathbb{Q}$ using $\mathbf{q}_t = \mathrm{Vec}(\log(\mathbf{Q}_t))$

2: Compute the soft-assignment of vector $\mathbf{q}_t$, $1 \leq t \leq p$ to all the components of GMM using (12).

3: **for** $i = 1 \to K$ **do**

4:     Compute i-th Fisher score,
$\mathcal{G}_i = \frac{1}{p\sqrt{\omega_i}} \sum_{t=1}^p \gamma_i(\mathbf{q}_t)\sigma_i^{-1}(\mathbf{q}_t - \mu_i)$, where $\Sigma_i = \sigma_i^2 \mathbb{I}$

5: **end for**

6: Concatenate the resulting Fisher scores to form the Fisher tensor, i.e., $FT(\mathbb{Q}) = [\mathcal{G}_1^T, \mathcal{G}_2^T, \ldots, \mathcal{G}_K^T]^T$

6. Empirical evaluations

In this section we compare and contrast the performance of the proposed BoRW and FT[3] against baseline methods like bag of SIFT words and several state-of-the-art approaches on ICPR [29] and SNPHEp-2 [12] datasets. The ICPR and SNPHEp-2 datasets are taken from two different laboratories, thus any method that achieves high performance on both can be considered more robust.

Before delving into experiments, we elaborate how a descriptive representation of HEp-2 cells can be attained by CovDs. To this end, from each cell image a set of CovDs is extracted and then passed to BoRW or FT to generate image signature. To generate CovDs from a cell image, a feature vector is assigned to each pixel in the image through using a Gabor filter-bank. More specifically, each pixel is described by

$$F_{x,y} = [I(x,y), x, y, |G_{0,0}(x,y)|, |G_{0,1}(x,y)|, \ldots, |G_{0,v}(x,y)|, |G_{1,0}(x,y)|, \ldots, |G_{u,v}(x,y)|],$$
(14)

where $I(x,y)$ is the intensity value at position x, y and $G_{u,v}(x,y)$ is the response of a 2D Gabor wavelet [55] centered at x, y with orientation u and scale v:

$$G_{u,v}(x,y) = \frac{k_v^2}{4\pi^2} \sum_{t,s} e^{-k_v^2/8\pi^2((x-s)^2 + (y-t)^2)} (e^{ik_v((x-t)\cos(\theta_u) + (y-s)\sin(\theta_u))} - e^{-2\pi^2}),$$

with $k_v = 1/\sqrt{2^{v-1}}$ and $\theta_u = \pi u/8$. Then, a set of overlapping blocks are extracted from the cell image and the CovD for each block is obtained. We emphasize that only foreground pixels will be considered in computing the CovD of each block. In FT framework, each FT is normalized in two steps. First, a *power normalization* is performed on each element of the FT based on the recommendation of Jegou et al. [28]. This is to avoid having

concentrated distribution around zero. The transfer function for *power normalization* is $y : \mathbb{R} \to \mathbb{R}$, $y(x) = \mathrm{sign}(x)\sqrt{|x|}$, where x is an element of FT and $|\cdot|$ denotes absolute value. This is followed by ℓ_2 normalization. To classify signatures for both BoRW and FT methods, Support Vector Machines (SVM) [50] were employed.

6.1. Material

ICPR dataset: The ICPR HEp-2 Cell Classification Contest (ICPR) dataset contains 1457 cells extracted from 28 specimen images and contains six different patterns: homogeneous, fine speckled, coarse speckled, nucleolar, centromere, and cytoplasmic. Each specimen image is taken by using a colored camera fitted on a fluorescence microscope (40-fold magnification lens) coupled with 50W mercury vapour lamp. The dataset is organized into one split training and testing (721 training images and 734 testing images). We will use the expressions *cell level* to consider a single cell without the surrounding part of the image and *image level* to consider the whole image on this dataset. All the segmentation masks of the images are provided in the dataset. Specialists manually segmented and annotated each cell. In particular, a biomedical engineer manually segmented the cells using a tablet PC. Subsequently, each image was verified by a medical doctor specializing in Immunology and with 11 years experience, who annotated information at both image and cell levels. At the image level, the specialist annotated if there are enough mitotic cells to ensure, as previously stated, the correct preparation of the sample, the fluorescence intensity and the staining pattern. See Fig. 4 for examples. Foggia et al. [33] analyses 28 methods attending in the first international contest on this dataset hosted by the 21st edition of the International Conference on Pattern Recognition (ICPR 2012).

SNPHEp-2 dataset: The SNP HEp-2 Cell (SNPHEp-2) dataset[4] has five patterns: centromere, coarse speckled, fine speckled, homogeneous, and nucleolar. There are 1884 cell images extracted from 40 specimen images. Each specimen image is captured by using a monochrome high dynamic range cooled microscopy camera fitted on a microscope with a plan-Apochromat 20x/0.8 objective lens and an LED illumination source. DAPI [56] image channel was used to automatically extract the cell image masks. The specimen images are divided evenly into training and testing sets (4 images per pattern in training and testing sets). In total there are 905 and 979 cell images extracted for training and testing. Five-fold validation is considered as test protocol here. Examples of this dataset are shown in Fig. 4.

6.2. Evaluation on ICPR dataset

We compare the proposed BoRW and FT approaches against three state-of-the-art cell classification systems: (1) Cordelli and Soda [9], (2) Strandmark et al. [30], and (3) Wiliem et al. [12] on the ICPR dataset. We implemented the descriptor that achieves the

---

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

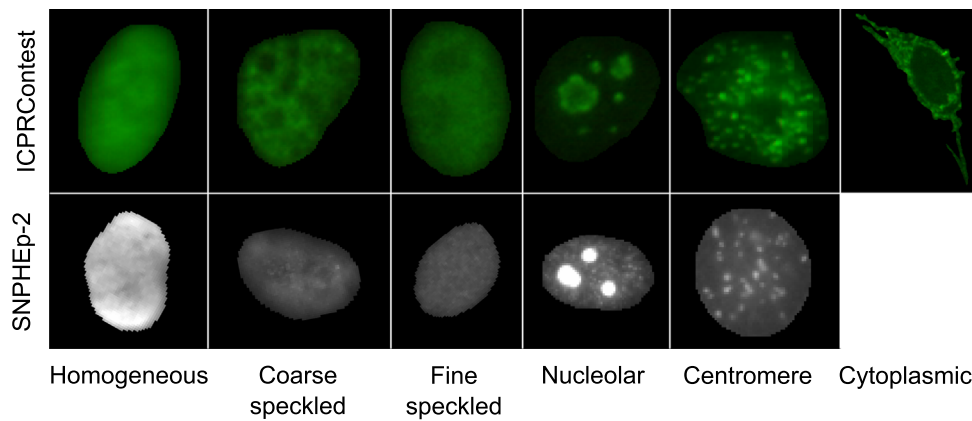[4] The SNPHEp-2 dataset is available for download at http://itee.uq.edu.au/~lovell/snphep2/.

Fig. 4. Sample images from ICPR [29] and SNPHEp-2 [12] datasets.

Table 1
Comparisons between the proposed approach to the BoRW and state-of-the-art methods on ICPR2012 dataset. CellCCR: Cell Correct Classification Rate (in %); ImageCCR: Image Correct Classification Rate (in %).

Method	CellCCR	ImageCCR
Cordelli and Soda [9]	49.9	N/A
Strandmark et al. [30]	48.0	N/A
Wiliem et al. [12]	55.3	N/A
Proposed BoRW ($k=1024$)	68.1	71.4
Proposed FT ($k=16$)	**70.2**	**78.6**

highest accuracy in [9] and denote the resulting system *Cordelli*. This resulted in a cell classification system that exploited image energy, mean, and entropy over the intensity and LBP channels as features. For the *Strandmark* system, we used the code provided in [30]. The *Strandmark* system utilized various image statistics (e.g., mean and standard deviation) and morphological features (e.g., number of objects, and area) along a random forest classifier for identification. As for the *Wiliem*, the BoW model in conjunction with the nearest convex Hull classifier that achieved the highest performance was used [12]. This resulted in Discrete Cosine Transform (DCT) as local patch descriptor and soft assignment as histogram encoding method. In addition the system used a dual region approach, i.e., two histograms, one from inner and one from outer region were used.

We present our results in both cell and image level classification performances. The image level accuracy is computed by considering the guessed classes for all cells of an specimen image (there are totally 28 specimen images in the dataset) and assigning the most frequently assigned class to the cells within that image. Table 1 compares the proposed BoRW and FT against *Cordelli*, *Strandmark*, and *Wiliem*. For FT, a GMM with 16 Gaussians achieved the best performance. We increased the number of Gaussians up to 128 but no significant improvement was observed. For BoRW, the vocabulary size is 1024. The effect of codebook size for both FT and BoRW will be addressed in Section 7. The table is self-explanatory. Both BoRW and FT outperforms state-of-the-art competitors significantly. The proposed FT obtained the highest accuracy amongst studied approaches. Fig. 5 shows the confusion matrices for cell and image level classification on ICPR.

Cell level confusion matrix shows that Centromere and Cytoplasmic cells are recognized much more reliably than other classes. That might be because the inherent textures in Centromere and Cytoplasmic cells are quite different from the remaining four classes. Also, the shape of Cytoplasmic cells could have a positive effect on the recognition rate of this class. Except

Cytoplasmic, over 10% of other cell types have a tendency to be misclassified as Centromere. Moreover, Centromere cells are mostly confused with Nucleolar (and vice versa), since both cell types have many dots on them with differences being number and size of dots.

Coarse Speckled are wrongly classified as Fine Speckled in around 20% of cases. We also note that the maximum misclassification happens between Homogeneous and Fine Speckled cells. Some representative examples of Homogeneous (Fine Speckled) cells misclassified as Fine Speckled (Homogeneous) are shown in Fig. 6. It is not surprising, because these two texture types are very similar in this dataset (see upper row in Fig. 4 for comparison). Our analysis over the whole dataset shows that the error rate for intermediate intensity cells is approximately 40%, while this number for positive intensity cells is approximately 20%. Most of the cells in Fig. 6 are of intermediate intensity cells.

Turning our attention to the image level confusion matrix, we note that FT could detect all Centromere, Nucleolar, and Cytoplasmic images accurately (i.e., the classification rate is the highest compared to other classes). Confusion between Homogenous and Fine Speckled cells and also Coarse Speckled and Fine Speckled cells results in three out of 14 image classification errors (see Fig. 5b).

The performance of proposed FT approach on ICPR dataset was also evaluated using Leave-One-Out validation protocol. More specifically, we created 28 splits of train and test images, where in each split, cell images of a particular specimen are considered as test images and the remaining images are used for training. Our proposed FT achieved 71.7% and 75% correct classification rate for cell and image level tests, respectively. This result along confusion matrices for cell and image level classification are shown in Figs. 7 and 8.

Similar to the previous experiment, Centromere and Cytoplasmic cells constitute the highest classification rates for cell and image, respectively. Moreover, Homogenous and Fine Speckled cells (and vice versa) are confused more often as compared to other cells. In the leave-one-out experiment, Nucleolar cells are more confused with Centromere and Cytoplasmic cells. This caused assigning two Nucleolar images to either of the classes. Also, confusion between two Speckled classes caused two image classification errors. Totally, FT was unable to correctly classify seven images, of which four are with intermediate intensity and three are with positive intensity.

6.3. Evaluation on SNP HEp-2

Similar to ICPR, we compared the proposed BoRW and FT methods against three cell classification systems: (1) Cordelli and Soda [9], (2) Strandmark et al. [30], and (3) Wiliem et al. [12] on SNP

a

	Ce	H	N	CS	FS	Cy
Ce	.90	.01	.09	.00	.00	.00
H	.11	.63	.03	.02	.20	.01
N	.12	.04	.71	.01	.04	.09
CS	.14	.00	.01	.65	.19	.01
FS	.11	.27	.04	.06	.49	.03
Cy	.04	.00	.02	.02	.00	.92

b

	Ce	H	N	CS	FS	Cy
Ce	1.0	.00	.00	.00	.00	.00
H	.00	.50	.00	.00	.50	.00
N	.00	.00	1.0	.00	.00	.00
CS	.00	.00	.00	.67	.33	.00
FS	.00	.50	.00	.00	.50	.00
Cy	.00	.00	.00	.00	.00	1.0

Fig. 5. Confusion matrices of the proposed FT on ICPR dataset. (a) Cell level confusion matrix, the average accuracy is 70.16%. (b) Image level confusion matrix, the average accuracy is 78.57%. Labels of row/column have the following meaning: Ce, centromere; H, homogenous; N, nucleolar; CS, coarse speckled; FS, fine speckled, and Cy, cytoplasmic.
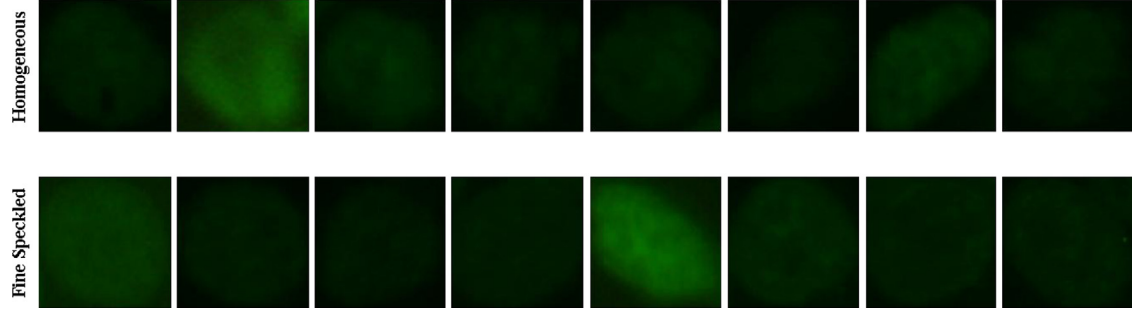
Fig. 6. Sample errors of our proposed FT method. Up: Homogeneous cells classified as Fine Speckled. Down: Fine Speckled cells classified as Homogeneous.

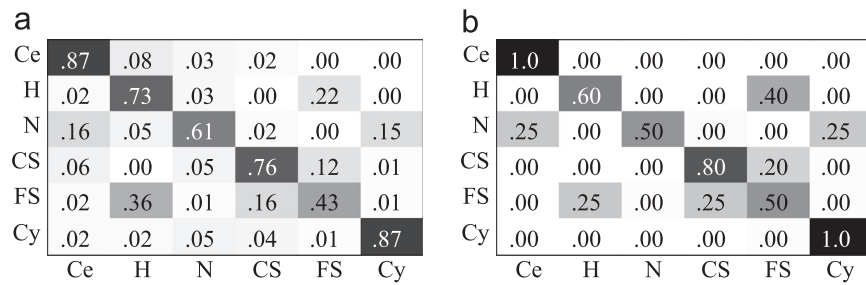| Image ID	Image class	Cells assigned to Centromere		Cells assigned to Homogeneous		Cells assigned to Nucleolar		Cells assigned to Coarse Speckled		Cells assigned to Fine Speckled		Cells assigned to Cytoplasmic	
		Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
3	Centromere	89	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
7	Centromere	47	83.9%	0	0.0%	9	16.1%	0	0.0%	0	0.0%	0	0.0%
13	Centromere	42	91.3%	0	0.0%	0	0.0%	4	8.7%	0	0.0%	0	0.0%
14	Centromere	40	63.5%	23	36.5%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
16	Centromere	36	94.7%	0	0.0%	1	2.6%	0	0.0%	1	2.6%	0	0.0%
19	Centromere	57	87.7%	6	9.2%	0	0.0%	2	3.1%	0	0.0%	0	0.0%
1	Homogeneous	0	0.0%	57	93.4%	0	0.0%	0	0.0%	4	6.6%	0	0.0%
5	Homogeneous	1	2.1%	16	34.0%	2	4.3%	0	0.0%	28	59.6%	0	0.0%
18	Homogeneous	0	0.0%	42	100.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
21	Homogeneous	5	8.2%	22	36.1%	1	1.6%	0	0.0%	33	54.1%	0	0.0%
22	Homogeneous	0	0.0%	104	87.4%	6	5.0%	0	0.0%	8	6.7%	1	0.8%
4	Nucleolar	9	13.6%	6	9.1%	21	31.8%	0	0.0%	0	0.0%	30	45.5%
8	Nucleolar	28	50.0%	2	3.6%	17	30.4%	3	5.4%	0	0.0%	6	10.7%
20	Nucleolar	1	2.2%	0	0.0%	44	95.7%	1	2.2%	0	0.0%	0	0.0%
24	Nucleolar	1	1.4%	5	6.8%	66	90.4%	0	0.0%	1	1.4%	0	0.0%
6	Coarse Speckled	0	0.0%	0	0.0%	0	0.0%	66	97.1%	2	2.9%	0	0.0%
10	Coarse Speckled	4	12.1%	0	0.0%	0	0.0%	22	66.7%	7	21.2%	0	0.0%
11	Coarse Speckled	0	0.0%	0	0.0%	10	24.4%	30	73.2%	1	2.4%	0	0.0%
12	Coarse Speckled	8	16.3%	0	0.0%	0	0.0%	36	73.5%	2	4.1%	3	6.1%
17	Coarse Speckled	1	5.3%	0	0.0%	0	0.0%	5	26.3%	13	68.4%	0	0.0%
2	Fine Speckled	0	0.0%	16	33.3%	0	0.0%	7	14.6%	25	52.1%	0	0.0%
9	Fine Speckled	0	0.0%	32	69.6%	0	0.0%	8	17.4%	6	13.0%	0	0.0%
15	Fine Speckled	4	6.3%	18	28.6%	3	4.8%	19	30.2%	17	27.0%	2	3.2%
23	Fine Speckled	1	2.0%	9	17.6%	0	0.0%	0	0.0%	41	80.4%	0	0.0%
25	Cytoplasmic	2	8.3%	2	8.3%	4	16.7%	2	8.3%	1	4.2%	13	54.2%
26	Cytoplasmic	0	0.0%	0	0.0%	1	2.9%	0	0.0%	0	0.0%	33	97.1%
27	Cytoplasmic	0	0.0%	0	0.0%	0	0.0%	2	5.3%	0	0.0%	36	94.7%
28	Cytoplasmic	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	13	100.0%

Fig. 7. True class for each cell image, number, and percentage of assignments to all of the existing classes on ICPR dataset.

HEp-2 dataset. Here, we only considered the cell level classification accuracy and report the results by averaging over five available splits of data. As shown in Table 2, the proposed approaches significantly outperform the state-of-the-art systems proposed by Cordelli and Strandmark on SNP HEp-2 dataset. The FT performance is to some extent close to that of Wiliem [12]. However, we note that Wiliem's system exploits the prior knowledge of image quality in its decision while our system does not. More specifically, in [12], two separate classifiers (a high-resolution and a low-resolution) are employed to classify a cell image. Based on the prior knowledge of image quality, only one of the classifiers will be invoked. On the contrary, our system does not require such prior knowledge for its decision and still achieves a decent accuracy.

We also show the average confusion matrix of five splits in Fig. 9. This figure shows that Homogeneous cells are detected more reliably as a result of their unique and different appearance

a

Ce	.87	.08	.03	.02	.00	.00
H	.02	.73	.03	.00	.22	.00
N	.16	.05	.61	.02	.00	.15
CS	.06	.00	.05	.76	.12	.01
FS	.02	.36	.01	.16	.43	.01
Cy	.02	.02	.05	.04	.01	.87
	Ce	H	N	CS	FS	Cy

b

Ce	1.0	.00	.00	.00	.00	.00
H	.00	.60	.00	.00	.40	.00
N	.25	.00	.50	.00	.00	.25
CS	.00	.00	.00	.80	.20	.00
FS	.00	.25	.00	.25	.50	.00
Cy	.00	.00	.00	.00	.00	1.0
	Ce	H	N	CS	FS	Cy

Fig. 8. Confusion matrices of the proposed FT on ICPR dataset using Leave-One-Out validation protocol. (a) Cell level confusion matrix, the average accuracy is 71.7%. (b) Image level confusion matrix, the average accuracy is 75.0%. Abbreviations are the same as used in Fig. 5.

Table 2
Comparisons between the proposed BoRW and FT approaches against the state-of-the-art methods on SNP HEp-2 dataset. CellCCR: Cell Correct Classification Rate (in %).

Method	CellCCR
Cordelli and Soda [9]	42.6
Strandmark et al. [30]	56.8
Wiliem et al. [12]	**80.6**
Proposed BoRW ($k=1024$)	67.6
Proposed FT ($k=16$)	74.7

Ce	.70	.00	.03	.27	.00
H	.00	.90	.00	.00	.10
N	.10	.00	.81	.08	.01
CS	.26	.00	.14	.57	.03
FS	.01	.17	.06	.01	.75
	Ce	H	N	CS	FS

Fig. 9. Cell level confusion matrix of the proposed FT on SNP HEp-2 dataset computed by averaging over the splits. Abbreviations are the same as used in Fig. 5.

in the dataset (see Fig. 4). However, similar to ICPR dataset the confusion between Homogeneous and Fine Speckled (and vice versa) is still notable. Around 25% of Centromere and Coarse Speckled cells are confused with each other which is the highest confusion value in this dataset.

7. Further discussions

In Section 6, we assessed the performance of the proposed BoRW and FT methods on two HEp-2 cell datasets. This section is dedicated to answer some basic questions about the proposed approaches. First, we compare the performance of proposed CovD descriptor in conjunction with BoRW and FT against popular Euclidean BoW models. This is to show the effectiveness of our proposed descriptor. We also study the effect of codebook size on recognition accuracy. Later, we study the effect of inter-laboratory imagery on the proposed approaches. More specifically, we evaluate the performance of our proposed BoRW and FT for various setups where training and test data do not necessarily come from the same source. Next, we assess the performance of BoRW and FT for a general texture classification task using KTH-TIPS [31] dataset. This is mainly driven by the fact that the proposed BoRW and FT can be seen more generally as frameworks for object recognition. Finally, we discuss the computational complexity of the proposed algorithms.
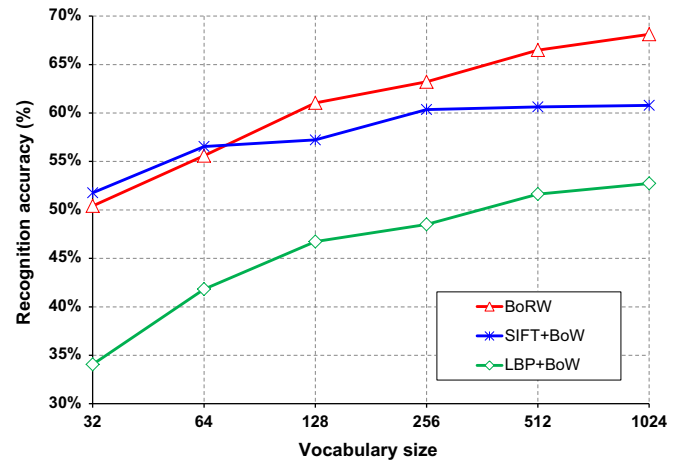
Fig. 10. Performance comparison of various block level representations on the ICPR.

7.1. Efficiency of CovD and effect of Codebook size

We analyze the performance of BoRW and the CovD descriptors for identification of HEp-2 cells in this subsection. To this end, we compare the performance of BoRW against two popular BoW methods, namely SIFT-BoW and LBP-BoW. The Scale Invariant Feature Transform (SIFT) [57] and Local Binary Pattern (LBP) [58] are powerful and robust features that have been successfully utilized in many computer vision applications. Fig. 10 compares the proposed BoRW against SIFT-BoW and LBP-BoW for various vocabulary sizes. Except for very small dictionaries, BoRW outperforms the SIFT-BoW and LBP-BoW significantly. The difference between BoRW and SIFT-BoW exceeds eight percentage points for vocabulary size of 1024. This clearly justifies our choice of CovD for identification of HEp-2 cells. We note that the superiority of CovDs for texture classification has been also observed in several recent studies [19,23,24].

7.2. Inter laboratory evaluations

In this subsection, we assess the robustness of the proposed methods against different staining patterns for cell classification. To this end, three types of experiments were conducted. First, we mixed the train and test sets of both ICPR and SNP HEp-2 datasets. This resulted in 1626 (721 cells from ICPR and 905 cells from SNP HEp-2) train images and 1713 (734 cells from ICPR and 979 cells from SNP HEp-2) test images in the combined dataset. Next, we delve into more challenging settings and analyze the performance of BoRW and FT where models are trained on one dataset and used in another one. More specifically, for both BoRW and FT, dictionaries were trained on the training set of either ICPR or SNP HEp-2 datasets and then used to extract signature in the other dataset

and referred to as "ICPR→SNP HEp−2" and "SNP HEp−2→ICPR", respectively. The last two tests can be understood as the domain shift problem [59].

The results for the three aforementioned experiments are shown in Table 3. For the domain shift experiments, unlike the approach proposed by Wiliem et al. [12] (which is a BoW model) and both BoRW and FT, there is no explicit modeling phase in Cordelli's [9] and Strandmark's [30] and hence the performances are not provided here. Moreover, we note that cytoplasmatic cells were excluded in testing on ICPR as there is no such cells in SNP HEp-2 dataset. We also note that the proposed BoRW method gives a decent performance in all the experiments.

Table 3 shows that the proposed FT method significantly outperforms all the competitors. For example, FT outperforms the closest competitor on ICPR+SNP HEp-2 by more than 2% percentage points. A more notable example is the SNP HEp−2→ICPR experiment, where the difference between FT and the Wiliem's method exceeds 19% percentage points. This is clearly an indication of robustness against different staining patterns, imaging conditions, and scales. Furthermore, in SNP HEp−2→ICPR, the performance of our methods dropped by around 10%. Excluding cytoplasmic cells (which can be classified easily) from the dataset could be a factor to this decrease of course. Last, but not least, the proposed BoRW method performs steady and equally well in all conducted experiments.

7.3. Texture classification

To show the capabilities of the proposed methods as general frameworks for object recognition, we performed an extra experiment using the KTH-TIPS [31] dataset. KTH-TIPS [31] (KTH Textures under varying Illumination, Pose, and Scale) dataset contains 810 texture images from 10 different classes (81 sample per class) captured at nine scales (relative scale changes from 0.5 to 2), three different illumination directions, and three different poses. Images in the dataset have no obvious rotation. Examples of all classes in three scales are shown in Fig. 11. It is apparent that textures in some classes look very similar to images of other classes at different scales.

We followed the protocol devised in [60–62] and considered half of images per each class for training and the remaining ones for testing. The performance is reported by averaging over 100 runs. We compare the performance of BoRW and FT against three state-of-the-art texture recognition systems [60–62].

Zhang et al. [60] propose a system where images are represented as distributions of extracted features from a set of keypoint locations. Three types of feature descriptors (SPIN [63], RIFT [63], and SIFT [35]) and a kernel SVM classifier with two measures (χ^2 and Earth Mover's distance) for comparing image distributions are used. In [61] texture histograms are obtained from a visual vocabulary of Basic Image Features [64] computed at every pixel at four octave-separated scales. Nearest neighbor classifier is used for classifying a query texture. In [62], Liu et al. propose to extract a set of random measurements from sorted pixel differences in local image patches. In order to obtain a global rotation invariance, local rotation invariant random features are embedded into a bag-of-words model. Nonlinear kernel SVM with χ^2 distance measure is then used for classification.

Table 4 compares the proposed BoRW and FT methods against Zhang et al. [60], Crosier and Griffin [61], and Liu et al. [62] approaches. We used the same features and classifiers as utilized in cell classification experiments. As can be seen in the table, our proposed FT could outperform all the three competitors. Furthermore, our method avoids complex design steps and expensive computational cost involved in the other methods. On average, the

Table 4
Comparisons between the proposed BoRW and FT approaches against the state-of-the-art methods on KTH-TIPS dataset. TextureCCR: Texture Correct Classification Rate (in %).

Method	TextureCCR
Zhang et al. [60]	96.1
Crosier and Griffin [61]	98.5
Liu et al. [62]	99.1
Proposed BoRW ($k=1024$)	97.2
Proposed FT ($k=16$)	**99.3**

Table 3
Cell correct classification rate (in %) comparisons between the proposed BoRW and FT approaches against the state-of-the-art methods on inter laboratory experiments.

Method	ICPR+SNP HEp-2	ICPR→SNP HEp−2	SNP HEp−2→ICPR
Cordelli and Soda [9]	42.8	N/A	N/A
Strandmark et al. [30]	60.2	N/A	N/A
Wiliem et al. [12]	65.6	41.3	49.0
BoRW ($k=1024$)	63.6	59.7	64.9
FT ($k=16$)	**67.7**	**60.2**	**68.4**

Fig. 11. Sample textures from KTH-TIPS dataset, showing three scales with the same illumination directions and poses for all classes. Classes from left to right are Aluminium Foil, Brown Bread, Corduroy, Cotton, Cracker, Linen, Orange Peel, Sandpaper, Sponge, and Styrofoam.

proposed FT method misclassified three images. Our analysis shows that two classes contributed the most in the error for the FT method. Few images in class Linen were confused with Corduroy, Cotton, or Sandpaper. Also, very few confusions between Brown Bread and Sponge (and vice versa) were observed.

7.4. Computational complexity

The computational complexity of BoRW and FT signature computation is analyzed here. In both cases, covariance matrices need to be computed from a test image. The covariance matrices can be computed efficiently (i.e., in one pass over the image) via the integral images [15]. This results in $O(WHd^2)$ flops for computing a $d \times d$ covariance matrix from a $W \times H$ image.

Generating histograms in BoRW methods requires computing K distances using the Stein metric. This requires computation of determinants, which can be done rapidly via Cholesky factorization, at the cost of $O(d^3)$. Therefore, computing a K dimensional BoRW signature for one covariance matrix demands $O(Kd^3)$ computations. For FT, CovDs need to be mapped to log-Euclidean space first. Again, the matrix logarithm can be computed using Cholesky factorization with $O(d^3)$ flops. Computing the FT as explained in the main loop of Algorithm 3 can be done in quadratic time in N where $N = d(d+1)/2$. Putting altogether, the computational complexity of BoRW and FT methods for one $W \times H$ image is $O(WHd^2 + Kd^3)$ and $O(WHd^2 + (2+K)d^3 + \frac{1}{2}d^4)$, respectively. We note that though the complexity of FT is of order 4, for small and medium size CovDs, the complexity is dictated by value K. Therefore, the FT method is an order of magnitude faster in our case since K is much smaller than that of BoRW method. For example, by averaging over 10 runs, computing BoRW signatures with $K=1024$ over 1000 images took 4647 s on a 3.4 GHz CPU, while the same test just needed 114 seconds for FT method with $K=16$.

8. Main findings and future directions

In this paper, we proposed a bridge between Bag of Word (BoW) models and a special class of non-Euclidean spaces, the space of Symmetric Positive Definite (SPD) matrices. To this end, we devised an intrinsic extension of conventional BoW using Riemannian geometry of SPD matrices. This was followed by introducing a Riemannian version of Fisher vectors [26] and utilizing it for analyzing and classifying Human Epithelial Type 2 Cell from indirect immunoFluoresence images.

Thorough experiments on two HEp-2 cell datasets, namely ICPRContest and SNP HEp-2, indicate that the proposed approach obtains notable improvements in discrimination accuracy in comparison to Euclidean and state-of-the-art methods such as the ones proposed by Cordelli and Soda [9], Strandmark et al. [30], and Wiliem et al. [12]. We further investigated the performance of our method on classifying textures from KTH-TIPS dataset against several state-of-the-art methods, including Zhang et al. [60], Crosier and Griffin [61], and Liu et al. [62].

The main ingredient of our proposal is a diffeomorphism that embeds Riemannian manifold of SPD matrices into a Euclidean space. While several studies [44,52–54] demonstrate the benefit of such embedding, the true geometry of space is not utilized into its full capacity. As such, future avenues of research include exploring a fully intrinsic formulation for the proposed log-Euclidean Fisher vector method. We believe that intrinsic extension of this work will intercept with intrinsic probabilistic models on Riemannian manifolds [17]. Moreover, we are keen to explore the efficiency of our proposal on generic object recognition tasks and assess and contrast its performance against Euclidean counterparts.

Conflict of interest statement

None declared.

References

[1] P.L. Meroni, P.H. Schur, ANA screening: an old test with new recommendations, Ann. Rheum. Dis. 69 (8) (2010) 1420–1422.
[2] A.S. Wiik, M. Høier-Madsen, J. Forslid, P. Charles, J. Meyrowitsch, Antinuclear antibodies: a contemporary nomenclature using HEp-2 cells, J. Autoimmun. 35 (2010) 276–290.
[3] P. Perner, H. Perner, B. Müller, Mining knowledge for HEp-2 cell image classification, Artif. Intell. Med. 26 (2002) 161–173.
[4] P. Soda, G. Iannello, Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis, IEEE Trans. Inf. Technol. Biomed. 13 (3) (2009) 322–329.
[5] R. Hiemann, N. Hilger, J. Michel, J. Nitschke, A. Bohm, U. Anderer, M. Weigert, U. Sack, Automatic analysis of immunofluorescence patterns of HEp-2 cells, Ann. N. Y. Acad. Sci. 1109 (1) (2007) 358–371.
[6] S.D. Cataldo, A. Bottino, E. Ficarra, E. Macii, Applying textural features to the classification of hep-2 cell patterns in IIF images, in: International Conference on Pattern Recognition (ICPR), 2012, pp. 3349–3352.
[7] P. Elbischger, S. Geerts, K. Sander, G. Ziervogel-Lukas, P. Sinah, Algorithmic framework for HEp-2 fluorescence pattern classification to aid auto-immune diseases diagnosis, in: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 562–565.
[8] T. Hsieh, Y. Huang, C. Chung, Y. Huang, HEp-2 cell classification in indirect immunofluorescence images, in: International Conference on Information, Communications and Signal Processing, 2009, pp. 1–4.
[9] E. Cordelli, P. Soda, Color to grayscale staining pattern representation in IIF, in: International Symposium on Computer-Based Medical Systems, 2011, pp. 1–6.
[10] I. Theodorakopoulos, D. Kastaniotis, G. Economou, S. Fotopoulos, Hep-2 cells classification via fusion of morphological and textural features, in: IEEE International Conference on Bioinformatics and Bioengineering (BIBE), 2012, pp. 689–694.
[11] G. Thibault, J. Angulo, Efficient statistical/morphological cell texture characterization and classification, in: International Conference on Pattern Recognition (ICPR), 2012, pp. 2440–2443.
[12] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, B.C. Lovell, Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors, in: IEEE Workshop on Applications of Computer Vision (WACV), 2013, pp. 95–102.
[13] G. Csurka, C. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in: Workshop on Statistical Learning in Computer Vision, ECCV, vol. 1, 2004, p. 22.
[14] O. Tuzel, F. Porikli, P. Meer, Region covariance: a fast descriptor for detection and classification, in: Proceedings of European Conference on Computer Vision (ECCV), vol. 3952, 2006, pp. 589–600.
[15] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on Riemannian manifolds, IEEE Trans. Pattern Anal. Mach. Intell. 30 (10) (2008) 1713–1727.
[16] F. Porikli, O. Tuzel, P. Meer, Covariance tracking using model update based on Lie algebra, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 728–735.
[17] X. Pennec, Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements, J. Math Imag. Vis. 25 (1) (2006) 127–154.
[18] Y. Pang, Y. Yuan, X. Li, Gabor-based region covariance matrices for face recognition, IEEE Trans. Circuits Syst. Video Technol. 18 (7) (2008) 989–993.
[19] A. Sanin, C. Sanderson, M.T. Harandi, B.C. Lovell, Spatio-temporal covariance descriptors for action and gesture recognition, in: IEEE Workshop on the Applications of Computer Vision (WACV), 2013.
[20] A. Goh, R. Vidal, Clustering and dimensionality reduction on Riemannian manifolds, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–7.
[21] R. Subbarao, P. Meer, Nonlinear mean shift over Riemannian manifolds, Int. J. Comput. Vision 84 (1) (2009) 1–20.

[22] S. Sra, A. Cherian, Generalized dictionary learning for symmetric positive definite matrices with application to nearest neighbor retrieval, in: European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD), Springer–Verlag, 2011, pp. 318–332.

[23] M.T. Harandi, C. Sanderson, A. Wiliem, B.C. Lovell, Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures, in: IEEE Workshop on the Applications of Computer Vision (WACV), 2012, pp. 433–439.

[24] M.T. Harandi, C. Sanderson, R. Hartley, B.C. Lovell, Sparse coding and dictionary learning for symmetric positive definite matrices: a kernel approach, in: Proceedings of European Conference on Computer Vision (ECCV), Springer, 2012, pp. 216–229.

[25] M.T. Harandi, C. Sanderson, S. Shirazi, B.C. Lovell, Kernel analysis on Grassmann manifolds for action recognition, Pattern Recognition Lett. 34 (15) (2013) 1906–1915.

[26] F. Perronnin, C. Dance, Fisher kernels on visual vocabularies for image categorization, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.

[27] F. Perronnin, J. Sánchez, T. Mensink, Improving the Fisher kernel for large-scale image classification, Proceedings of European Conference on Computer Vision (ECCV) (2010) 143–156.

[28] H. Jégou, F. Perronnin, M. Douze, C. Schmid, et al., Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1704–1716.

[29] P. Foggia, G. Percannella, P. Soda, M. Vento, Early experiences in mitotic cells recognition on hep-2 slides, in: International Symposium on Computer-Based Medical Systems (CBMS), 2010, pp. 38–43.

[30] P. Strandmark, J. Ulén, F. Kahl, Hep-2 staining pattern classification, in: International Conference on Pattern Recognition (ICPR), 2012, pp. 33–36.

[31] E. Hayman, B. Caputo, M. Fritz, J.-O. Eklundh, On the significance of real-world conditions for material classification, in: Proceedings of European Conference on Computer Vision (ECCV), Springer, 2004, pp. 253–266.

[32] Y. Yang, A. Wiliem, A. Alavi, P. Hobson, Classification of human epithelial type 2 cell images using independent component analysis, in: IEEE International Conference on Image Processing (ICIP), 2013.

[33] P. Foggia, G. Percannella, P. Soda, M. Vento, Benchmarking HEp-2 cells classification methods, IEEE Trans. Med. Imaging 32 (2013) 1878–1889.

[34] J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, Int. J. Comput. Vis.) 73 (2) (2007) 213–238.

[35] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[36] T. Tommasi, F. Orabona, B. Caputo, Discriminative cue integration for medical image annotation, Pattern Recognition Lett. 29 (15) (2008) 1996–2002.

[37] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2006, pp. 2169–2178.

[38] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.

[39] O. Boiman, E. Shechtman, M. Irani, In defense of nearest-neighbor based image classification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[40] I. Kolár, J. Slovak, P. Michor, Natural Operations in Differential Geometry, Springer, 1999.

[41] R. Bhatia, Positive Definite Matrices, Princeton University Press, 2007.

[42] Y.M. Lui, Advances in matrix manifolds for computer vision, Image Vis. Comput. Elsevier, 30 (6) (2012) 380–388.

[43] S. Sra, A new metric on the manifold of kernel matrices with application to matrix geometric means, Proceedings of Advances in Neural Information Processing Systems (NIPS), (2012) 144–152.

[44] V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 29 (1) (2007) 328–347.

[45] A.C. Thompson, On certain contraction mappings in a partially ordered vector space, Proceedings of the American Mathematical Society 14 (1963) 438–443.

[46] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos, Jensen-Bregman Logdet divergence with application to efficient similarity search for covariance matrices, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2012) 2161–2174.

[47] T.S. Jaakkola, D. Haussler, et al., Exploiting generative models in discriminative classifiers, Proceedings of Advances in Neural Information Processing Systems (NIPS) (1999) 487–493.

[48] J. Krapac, J. Verbeek, F. Jurie, Modeling spatial layout with fisher vectors for image categorization, in: Proceedings of International Conference on Computer Vision (ICCV), 2011, pp. 1487–1494.

[49] S.-i. Amari, H. Nagaoka, Methods of Information Geometry, vol. 191, American Mathematical Society, 2007.

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

[51] G. Cheng, B. Vemuri, A novel dynamic system in the space of SPD matrices with applications to appearance tracking, SIAM J. Imag. Sci. 6 (1) (2013) 592–615.

[52] R. Caseiro, P. Martins, J.F. Henriques, J. Batista, A nonparametric Riemannian framework with application to foreground segmentation, Pattern Recognition 45 (11) (2012) 3997–4017.

[53] W. Hu, X. Li, W. Luo, X. Zhang, S. Maybank, Z. Zhang, Single and multiple object tracking using log-Euclidean Riemannian subspace and block-division appearance model, IEEE Trans. Pattern Anal. Mach. Intell. 34 (12) (2012) 2420–2440.

[54] K. Guo, P. Ishwar, J. Konrad, Action recognition from video using feature covariance matrices, IEEE Trans. Image Process. 22 (6) (2013) 2479–2494.

[55] T.S. Lee, Image representation using 2d Gabor wavelets, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 959–971.

[56] R. Hiemann, T. Büttner, T. Krieger, D. Roggenbuck, U. Sack, K. Conrad, Challenges of automated screening and differentiation of non-organ specific autoantibodies on hep-2 cells, Autoimmun. Rev. 9 (1) (2009) 17–22.

[57] D.G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of International Conference on Computer Vision (ICCV), vol. 2, 1999, pp. 1150–1157.

[58] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.

[59] B. Kulis, K. Saenko, T. Darrell, What you saw is not what you get: domain adaptation using asymmetric kernel transforms, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1785–1792.

[60] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, Int. J. Comput. Vis. 73 (2) (2007) 213–238.

[61] M. Crosier, L.D. Griffin, Using basic image features for texture classification, Int. J. Comput. Vis. 88 (3) (2010) 447–460.

[62] L. Liu, P. Fieguth, D. Clausi, G. Kuang, Sorted random projections for robust rotation-invariant texture classification, Pattern Recognition 45 (6) (2012) 2405–2418.

[63] S. Lazebnik, C. Schmid, J. Ponce, A sparse texture representation using local affine regions, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1265–1278.

[64] L.D. Griffin, M. Lillholm, Feature category systems for 2nd order local image structure induced by natural image statistics and otherwise, in: Electronic Imaging, International Society for Optics and Photonics, 2007, pp. 649209–649209–11.

Masoud Faraki received his B.Sc. degree in Computer Engineering (Software) in 2006 and his M.Sc. degree in Computer Engineering (Artificial Intelligence) in 2009, both form Isfahan University of Technology, Iran. He is currently doing an internship at the University of Queensland (Australia). His research interests are Machine Learning and Computer Vision.

Mehrtash T. Harandi received his Ph.D. from University of Tehran, Iran, in 2009. In November 2009, he joined NICTA and the University of Queensland (Australia) as a Research Scientist. He has previously worked on active learning and object recognition at the Institute for Research in Fundamental Sciences. His current research is mainly focused on Riemannian geometry and its applications in machine learning and computer vision.

Arnold Wiliem is a research fellow at the University of Queensland. He received his Ph.D. in 2010 from Queensland University of Technology and his bachelor of Computer Science in 2006 from University of Indonesia. He has previous worked on automated medical image analysis at NICTA for 2 years before he joined the University of Queensland to pursue his research in digital pathology domain. His current research interests are in the areas of machine learning, pattern recognition, computer vision. He is a member of IEEE.

Brian C. Lovell is a professor at the University of Queensland, Australia. He was a research Leader at NICTA from 2006 to 2012. From 1983 to 1986, he worked in Saudi Arabia and Egypt with Schlumberger SA. He then returned to the University of Queensland to pursue his Ph.D. While studying for his Ph.D., he was offered a lecturing position in 1989. In 2000, Professor Lovell was visiting professor at the MIT Media Lab, Michigan State University, and the University of Alberta. In 2004 he was appointed as Director of Engineering in the School of ITEE at the University of Queensland. He is a Past-President of the International Association for Pattern Recognition (IAPR).