# Texture image classification with Riemannian fisher vectors

Ioana Ilea, Lionel Bombrun, Christian Germain, Romulus Terebes, Monica Borda, Yannick Berthoumieu

# TEXTURE IMAGE CLASSIFICATION WITH RIEMANNIAN FISHER VECTORS

*Ioana Ilea*[1,2], *Lionel Bombrun*[1], *Christian Germain*[1], *Romulus Terebes*[2], *Monica Borda*[2] *and Yannick Berthoumieu*[1]

[1]: Université de Bordeaux, Laboratoire IMS, Groupe Signal et Image.
{*ioana.ilea, lionel.bombrun, christian.germain, yannick.berthoumieu* }*@ims-bordeaux.fr*
[2]: Technical University of Cluj-Napoca. {*romulus.terebes, monica.borda* }*@com.utcluj.ro*

## ABSTRACT

This paper introduces a generalization of the Fisher vectors to the Riemannian manifold. The proposed descriptors, called Riemannian Fisher vectors, are defined first, based on the mixture model of Riemannian Gaussian distributions. Next, their expressions are derived and they are applied in the context of texture image classification. The results are compared to those given by the recently proposed algorithms, bag of Riemannian words and R-VLAD. In addition, the most discriminant Riemannian Fisher vectors are identified.

***Index Terms*—** Riemannian Fisher vectors, bag of words, Riemannian Gaussian distributions, classification, covariance matrix.

## 1. INTRODUCTION

Bag of words, Fisher vectors, or vectors of locally aggregated descriptors represent some of the most frequently used local models in order to capture the information lying in signals [1], images [2] or videos [3]. These descriptors have multiple advantages. First, the obtained information can be used in a wide variety of applications like classification [2] and categorization [4], text [5] and image [6] retrieval, action and face recognition [7], etc. Second, combined with powerful local feature descriptors such as SIFT, they are robust to transformations like scaling, translation, or occlusion [7].

The *bag of words* (BoW) model has been used for text retrieval and categorization [5, 8] and then extended to visual categorization [9]. This method is based on the construction of a codebook, or a dictionary, that contains the most significant features in the dataset. Generally, the elements in the codebook, or the words, are the clusters' centroids obtained by using the conventional k-means clustering algorithm. Next, for each element in the dataset, its signature is determined by computing the histogram of the number of occurrences of each word in its structure. To improve the performance of BoW, which counts only the number of local descriptors assigned to each Voronoi region, *Fisher vectors* (FV) have been introduced by including other statistics, such as the mean and variance of local descriptors.

FV are descriptors based on Fisher kernels [1], representing methods for measuring if samples are correctly fitted by some given models. By using FV, a sample is characterized by the gradient vector of the probability density function that models it, classically a Gaussian mixture model (GMM) [4]. In practice, the probability density function is replaced by the log-likelihood and, as mentioned in [4], its gradient describes the direction in which parameters should be modified to best fit the data. The derivatives with respect to the model's parameters are computed and concatenated to obtain the FV.

The *vectors of locally aggregated descriptors* (VLAD) represent a simplification of the Fisher kernel [10], based on the definition of a codebook. In the computation process, first of all, the dictionary has to be built. For this reason, the dataset is partitioned by using a clustering algorithm and the cluster centroids represent the codebook elements. Next, each element in the dataset is associated to the closest cluster. Further on, for each cluster a vector is computed, containing the differences between the cluster's centroid and each element in that cluster. In the end, the sum of differences concerning each cluster is computed and the final VLAD feature vector is given by the concatenation of all the previously obtained sums. In other way, the VLAD descriptors can be obtained starting from FV, by taking into consideration only the derivatives with respect to the means of the GMM. Note also that the homoscedasticity assumption and the hard assignment scheme are required to obtain VLAD features [7, 10].

Those three approaches have been widely used for many applications involving non-parametric features. Recently BoW and VLAD have been extended to the case where each feature is a point on a Riemannian manifold. This is for instance the case where local descriptors are covariance matrices. This includes many different applications in image processing, like classification [11, 12, 13], image segmentation [14], object detection [15, 16], etc. In [3] and [17], the BoW approach has been extended to the so-called log-Euclidean bag of words (LE-BoW) and bag of Riemannian words (BoRW) models by considering respectively the log-Euclidean and geodesic distance between two points on the manifold. In addition, the Riemannian version of the VLAD method (R-VLAD) has been developed in [7] and has shown superior classification performances, compared to the classic VLAD algorithm.

Until now, FV have not yet been generalized in the same manner to Riemannian manifold, due to the lack of probabilistic generative models suited for parametric descriptors. This represents the main contribution of this paper. The proposed Riemannian Fisher vectors (RFV) are a generalization of the FV for parametric descriptors based on the recent works on the definition of the Riemannian Gaussian distributions (RGDs) [18].

The paper is structured as follows. Section 2 recalls some elements on the RGD like its definition, the expression of mixtures of RGDs and the parameter's estimation procedure. Section 3 introduces the definition of the proposed RFV, their computation and their relation with R-VLAD. Section 4 presents an application of the proposed RFV to texture image classification. Conclusions and future works are finally reported in Section 5.

## 2. RIEMANNIAN GAUSSIAN DISTRIBUTIONS

Let $\mathbf{\Upsilon} = \{\mathbf{Y}_t\}_{t=1:T}$ be a set of $T$ independent and identically distributed (i.i.d.) samples according to a Riemannian Gaussian distribution of central value $\bar{\mathbf{Y}}$ and dispersion $\sigma$. The probability density function of the RGD with respect to the Riemannian volume element, in the space $\mathcal{P}_m$ of $m \times m$ real, symmetric and positive definite matrices, has been introduced in [18] as:

$$p(\mathbf{Y}_t|\bar{\mathbf{Y}}, \sigma) = \frac{1}{Z(\sigma)} \exp\Big\{ -\frac{d^2(\mathbf{Y}_t, \bar{\mathbf{Y}})}{2\sigma^2} \Big\}, \qquad (1)$$

where $Z(\sigma)$ is a normalization factor independent of the centroid $\bar{\mathbf{Y}}$ and $d(\cdot)$ is the Riemannian distance given by $d(\mathbf{Y}_1, \mathbf{Y}_2) = \left[\sum_i (\ln \lambda_i)^2\right]^{\frac{1}{2}}$, with $\lambda_i$, $i = 1, \dots, m$ being the eigenvalues of $\mathbf{Y}_2^{-1}\mathbf{Y}_1$.

Starting from (1), the probability density function for a mixture of $K$ RGDs can be defined as [18]:

$$p(\mathbf{Y}_t|\lambda) = \sum_{j=1}^{K} \varpi_j \, p(\mathbf{Y}_t|\bar{\mathbf{Y}}_j, \sigma_j), \qquad (2)$$

where $\lambda = \{(\varpi_j, \bar{\mathbf{Y}}_j, \sigma_j)_{1\leq j\leq K}\}$ is the parameter vector. $\varpi_j$ are positive weights, with $\sum_{j=1}^{K} \varpi_j = 1$ and $p(\mathbf{Y}_t|\bar{\mathbf{Y}}_j, \sigma_j)$ is given by (1).

Several approaches can be employed to estimate the parameters $\{\hat{\bar{\mathbf{Y}}}_j, \hat{\sigma}_j, \hat{\varpi}_j\}_{1\leq j\leq K}$ of the mixture of $K$ RGDs [12]. The simplest one implies the estimation of the centroids $\hat{\bar{\mathbf{Y}}}_j$, of clusters $c_j$, $j = 1, \dots, K$ by using the intrinsic k-means algorithm on a Riemannian manifold [7]. Thus, for each cluster $c_j$, the cost function

$$\varepsilon(\bar{\mathbf{Y}}_j) = \frac{1}{N_j} \sum_{n=1}^{N_j} d^2(\bar{\mathbf{Y}}_j, \mathbf{Y}_{j_n}) \qquad (3)$$

has to be minimized, where $\mathbf{Y}_{j_n}$ is the set of elements $\mathbf{Y}_j$ in cluster $c_j$, $n = 1, \dots, N_j$ and $N_j$ is the cardinal of $\mathbf{Y}_{j_n}$. The minimizer of the cost function defined in (3) is known to be the Riemannian centre of mass of this set. The interested reader is referred to [19] and [20] for an algorithm to compute the empirical Riemannian centre of mass. Next, for each cluster $c_j$, the estimated dispersion parameter $\hat{\sigma}_j$ is obtained as the solution of:

$$\sigma_j^3 \times \frac{d}{d\sigma_j} Z(\sigma_j) = \varepsilon(\hat{\bar{\mathbf{Y}}}_j). \qquad (4)$$

This latter is solved by a conventional Newton-Raphson algorithm [12]. Finally, the estimated weights $\hat{\varpi}_j$ are given by:

$$\hat{\varpi}_j = \frac{N_j}{\sum_{j=1}^{K} N_j}. \qquad (5)$$

All the elements recalled in this part are applied in the next section to the definition of the proposed Riemannian Fisher vectors.

## 3. RIEMANNIAN FISHER VECTORS

### 3.1. Definition

Let $\mathbf{\Upsilon} = \{\mathbf{Y}_t\}_{t=1:T}$ be a sample of $T$ i.i.d observations following a mixture of $K$ RGDs. Under the independence assumption, the probability density function of $\mathbf{\Upsilon}$ is given by:

$$p(\mathbf{\Upsilon}|\lambda) = \prod_{t=1}^{T} p(\mathbf{Y}_t|\lambda), \qquad (6)$$

where $\lambda = \{(\varpi_j, \bar{\mathbf{Y}}_j, \sigma_j)_{1\leq j\leq K}\}$ is the parameter vector and $p(\mathbf{Y}_t|\lambda)$ is the probability density function given in (2).

By using the Fisher kernels, the sample is characterized by its deviation from the model [2]. This deviation is measured by computing the Fisher score $U_{\mathbf{\Upsilon}}$ [1], that is the gradient $\nabla$ of the log-likelihood with respect to the model parameters $\lambda$:

$$U_{\mathbf{\Upsilon}} = \nabla_\lambda \log p(\mathbf{\Upsilon}|\lambda) = \nabla_\lambda \sum_{t=1}^{T} \log p(\mathbf{Y}_t|\lambda). \qquad (7)$$

As mentioned in [1], the gradient of the log-likelihood with respect to a parameter describes the contribution of that parameter to the generation of a particular observation. In practice, a large value for this derivative is equivalent to a large deviation from the model. Further on, that can be translated into the fact that the model does not correctly fit the data.

In the following, the derivatives for the mixture of RGDs, are given, knowing that $\gamma_i(\mathbf{Y}_t)$ is the probability that the observation $\mathbf{Y}_t$ is generated by the $i^{th}$ RGD and it is computed as:

$$\gamma_i(\mathbf{Y}_t) = \frac{\varpi_i \, p(\mathbf{Y}_t|\bar{\mathbf{Y}}_i, \sigma_i)}{\sum_{j=1}^{K} \varpi_j \, p(\mathbf{Y}_t|\bar{\mathbf{Y}}_j, \sigma_j)}. \qquad (8)$$

To determine the gradient with respect to the weight, we consider the procedure described in [2]. For that, the following parametrization is used in order to ensure the positivity and sum to one constraints of the weights:

$$\varpi_i = \frac{\exp(\alpha_i)}{\sum_{j=1}^{K} \exp(\alpha_j)}. \tag{9}$$

By taking into consideration all these observations, the derivatives with respect to the parameters in $\lambda$ can be obtained as:

$$\frac{\partial \log p(\boldsymbol{\Upsilon}|\lambda)}{\partial \bar{\mathbf{Y}}_i} = \sum_{t=1}^{T} \gamma_i(\mathbf{Y}_t)\, \sigma_i^{-2}\, \mathrm{Log}_{\bar{\mathbf{Y}}_i}(\mathbf{Y}_t), \tag{10}$$

$$\frac{\partial \log p(\boldsymbol{\Upsilon}|\lambda)}{\partial \sigma_i} = \sum_{t=1}^{T} \gamma_i(\mathbf{Y}_t) \left\{ -\frac{Z'(\sigma_i)}{Z(\sigma_i)} + \frac{d^2(\mathbf{Y}_t, \bar{\mathbf{Y}}_i)}{\sigma_i^3} \right\}, \tag{11}$$

$$\frac{\partial \log p(\boldsymbol{\Upsilon}|\lambda)}{\partial \alpha_i} = \sum_{t=1}^{T} \gamma_i(\mathbf{Y}_t)\, (1 - \varpi_i), \tag{12}$$

where $\mathrm{Log}_{\bar{\mathbf{Y}}_i}(\cdot)$ is the Riemannian logarithm mapping.

The vectorized representation of the derivatives in (10), (11) and (12) of the log-likelihood, with respect to the parameters in $\lambda$, gives the Riemannian Fisher vectors (RFV). In the end, by using the RFV, a sample is characterized by a feature vector containing some, or all the derivatives, having the maximum length given by the number of parameters in $\lambda$.

### 3.2. Relation with R-VLAD

As mentioned earlier in the introduction, VLAD features are a special case of FV. Therefore, R-VLAD can be viewed as a particular case of the proposed RFV. More precisely, R-VLAD is obtained by taking into consideration only the derivatives with respect to the central value $\bar{\mathbf{Y}}_i$ (see (10)). In addition, a hard assignment scheme is applied. Starting from the definition of the elements $v_i$ in the R-VLAD descriptor [7]:

$$v_i = \sum_{\mathbf{Y}_t \in c_i} \mathrm{Log}_{\bar{\mathbf{Y}}_i}(\mathbf{Y}_t), \tag{13}$$

with $\mathbf{Y}_t \in c_i$ being the elements $\mathbf{Y}_t$ assigned to the cluster $c_i$, $i = 1, \ldots, K$, the hard assignment implies that:

$$\gamma_i(\mathbf{Y}_t) = \begin{cases} 1, & \text{if } \mathbf{Y}_t \in c_i \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Moreover, the assumption of homoscedasticity is considered, that is $\sigma_i = \sigma\ , \forall i = 1, \ldots, K$. By considering these two assumptions, it is clear that (10) reduces to (13) hence confirming that RFV are a generalization of R-VLAD descriptors.

## 4. APPLICATION TO TEXTURE IMAGE CLASSIFICATION

This section introduces an application to texture image classification. The aim of this experiment is first to analyze the potential of the proposed RFV compared to the recently proposed bag of Riemannian words (BoRW) model [17] and R-VLAD [7]. The BoRW, RFV and R-VLAD are built based on region covariance descriptors [21] containing basic information, like image intensity and gradients. The experiment's purpose is not to find the best classification rates, but to compare the two methods starting from very simple descriptors. Second, the objective is to determine the RFV that are the most discriminant to retrieve the classes: the one associated to $\bar{\mathbf{Y}}_i$ (10), $\sigma_i$ (11) or $\alpha_i$ (12).

### 4.1. Databases

For this work, two texture databases are used: the VisTex [22] database and the Outex_TC000_13 [23] database. The *VisTex* database consists in 40 texture classes. Each class is composed of 64 images of size $64 \times 64$ pixels. The *Outex_TC000_13* database contains 68 texture classes, where each class is represented by a set of 20 images of size $128 \times 128$ pixels. For both databases, the feature extraction and classification steps are similar and are detailed in the next subsection.

### 4.2. Feature extraction and classification

For the classification procedure, the considered database is equally and randomly divided in order to obtain the training and the testing sets. For each image in the two sets, local descriptors have to be extracted first. In this experiment, the region covariance descriptors (RCovDs) are considered. In order to build the RCovD for an image $I$ of size $W \times H$, several characteristics are extracted for each pixel $(x, y) \in I$. Here, the image intensities $I(x, y)$ and the norms of the first and second order derivatives of $I(x, y)$ in both directions $x$ and $y$ are considered [21]. Thus, a vector $\mathbf{v}$ of 5 elements is obtained for each pixel having the spatial position $(x, y) \in I$:

$$\mathbf{v}(x, y) = \left[ I(x, y),\ \left|\frac{\partial I(x,y)}{\partial x}\right|,\ \left|\frac{\partial I(x,y)}{\partial y}\right|,\ \left|\frac{\partial^2 I(x,y)}{\partial x^2}\right|,\ \left|\frac{\partial^2 I(x,y)}{\partial y^2}\right| \right]^T. \tag{15}$$

For the two considered databases, the extracted RCovD are the estimated covariance matrices of vectors $\mathbf{v}(x, y)$ computed on a sliding patch of size $15 \times 15$ pixels. As an overlap of 8 pixels is considered for the patches, the VisTex and Outex databases are represented respectively by a set of 36 and 196 covariance matrices per texture class (of size $5 \times 5$). To speed-up the computation time, the fast covariance computation algorithm based on integral images presented in [21] has been implemented. In the end, each texture class is characterized by a set $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ of $N$ covariance matrices, that are elements in $\mathcal{P}_5$. Based on the patches in the training set, a codebook is created. For each class, the codewords are represented by the estimated parameters $\{\widehat{\bar{\mathbf{Y}}}_j, \hat{\sigma}_j, \hat{\varpi}_j\}_{1 \leq j \leq K}$ of the mixture of $K$ RGDs defined in (2). The estimation procedure is carried out here by using the intrinsic k-means algorithm (see Section 2). For this experiment, the number of modes $K$ is set to 3. In the end, the codebook is obtained by concatenating the previously extracted codewords.

Starting from the RCovDs and the learned codebook, the BoRW, RFV and R-VLAD local models are derived, as presented in the previous section. After their computation, a normalization stage is performed. In the RFV framework, the classical power and $\ell_2$ normalizations are applied [17]. The $\ell_2$ normalization has been proposed in [24] to minimize the influence of the background information on the image signature, while the power normalization corrects the independence assumption made on the patches [25]. The same normalization scheme is also applied for R-VLAD models. For the BoRW algorithm, only $\ell_2$ normalization is performed, as recommended in [3].

For the classification step, the SVM algorithm with Gaussian kernel is considered, knowing that the dispersion parameter of the Gaussian kernel is optimized by using a cross validation procedure on the training set.

### 4.3. Results

The classification performances in term of overall accuracy on the VisTex and Outex_TC000_13 databases are reported in Tables 1 and 2 respectively. Those results are displayed for 10 random partitions in training and testing sets. Columns homoscedasticity and prior correspond respectively to the homoscedasticity assumption and to the use of the weights $\varpi_i$ in the decision rule. If the homoscedasticity assumption is true, the dispersion parameter $\sigma_i$ is the same for all the clusters $c_i$. If the prior parameter is set to false, all the clusters have the same weight. Note that for the BoRW approach published in [17] and the R-VLAD presented in [7], the dispersion and weight parameters were not considered. Note also that for the proposed RFV, those two parameters are respectively set to "false" and "true", since both the dispersion and weight parameters are considered in the derivation of the RFV.

In this experiment, we also analyze the contribution of each parameter (weight, dispersion and centroid) to the classification accuracy. For example, the row "RFV : $\varpi$" indicates the classification results when only the derivatives with respect to the weights are considered to calculate the RFV (see (12)), ...

As observed in Tables 1 and 2, the proposed RFV outperforms the BoRW and R-VLAD approaches. A gain of 1 to 3% is observed for the VisTex database. Moreover, among the RFVs types, the most discriminant feature is obtained by combining the derivatives with respect to all three parameters: centroid, dispersion and weight (see (10), (11), (12)).

| **Method** | **Homoscedasticity** | **Prior** | **Overall accuracy** |
|---|---|---|---|
| BoRW | false | true | $87.22 \pm 1.19$ |
| BoRW | false | false | $87.51 \pm 0.92$ |
| BoRW [17] | true | false | $\mathbf{87.20 \pm 1.55}$ |
| BoRW | true | true | $76.67 \pm 2.35$ |
| RFV : $\varpi$ | false | true | $90.31 \pm 0.94$ |
| RFV : $\sigma$ | false | true | $81.42 \pm 1.12$ |
| RFV : $\bar{\mathbf{Y}}$ | false | true | $87.22 \pm 1.15$ |
| RFV : $\sigma, \varpi$ | false | true | $83.05 \pm 1.15$ |
| RFV : $\bar{\mathbf{Y}}, \varpi$ | false | true | $87.85 \pm 0.97$ |
| RFV : $\bar{\mathbf{Y}}, \sigma$ | false | true | $90.41 \pm 0.86$ |
| RFV : $\bar{\mathbf{Y}}, \sigma, \varpi$ | false | true | $\mathbf{90.43 \pm 0.84}$ |
| R-VLAD [7] | true | false | $\mathbf{87.94 \pm 0.58}$ |

**Table 1**. Classification results on the VisTex database.

| **Method** | **Homoscedasticity** | **Prior** | **Overall accuracy** |
|---|---|---|---|
| BoRW | false | true | $84.32 \pm 0.99$ |
| BoRW | false | false | $84.37 \pm 1.28$ |
| BoRW [17] | true | false | $\mathbf{84.43 \pm 1.23}$ |
| BoRW | true | true | $79.31 \pm 1.86$ |
| RFV : $\varpi$ | false | true | $84.94 \pm 1.12$ |
| RFV : $\sigma$ | false | true | $78.46 \pm 1.54$ |
| RFV : $\bar{\mathbf{Y}}$ | false | true | $83.94 \pm 0.90$ |
| RFV : $\sigma, \varpi$ | false | true | $80.38 \pm 1.80$ |
| RFV : $\bar{\mathbf{Y}}, \varpi$ | false | true | $84.26 \pm 0.75$ |
| RFV : $\bar{\mathbf{Y}}, \sigma$ | false | true | $84.32 \pm 1.19$ |
| RFV : $\bar{\mathbf{Y}}, \sigma, \varpi$ | false | true | $\mathbf{84.12 \pm 1.15}$ |
| R-VLAD [7] | true | false | $\mathbf{82.99 \pm 1.19}$ |

**Table 2**. Classification results on the Outex database.

## 5. CONCLUSION

In this paper, a new local model for image classification in the Riemannian space has been proposed. The introduced method, called Riemannian Fisher vectors, is a generalization of the so-called Fisher vectors, when the features are represented by parametric descriptors, like covariance matrices. The definition and the expression of RFV have been given, starting from the definition of the mixture of Riemannian Gaussian distributions. In addition, its relation with R-VLAD has been illustrated. In the end, the RFVs have been applied for texture image classification on the VisTex and Outex_TC000_13 databases. The results have been compared with those given by BoRW and R-VLAD, showing better classification rates for the same codebook. In addition, it has been observed that the most discriminant feature is obtained by combining the derivatives with respect to all parameters.

Further works on this subject will concern the extension of RFV to the recently proposed mixture of Riemannian Laplace distributions [26, 27].

## Acknowledgments

This study has been carried out in the frame of the Investments for the future Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02) of the French National Research Agency (ANR). It has also been supported by the French Foreign Affairs and International Development Ministry and by the Executive Agency for Higher Education, Research, Development and Innovation Funding Romania, under the projects 32619VL and PNII Capacitati 779/27.06.2014.

## 6. REFERENCES


[1] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *In Advances in Neural Information Processing Systems 11*. 1998, pp. 487–493, MIT Press.

[2] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the Fisher vector: Theory and practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.

[3] M. Faraki, M. Palhang, and C. Sanderson, "Log-euclidean bag of words for human action recognition," *Computer Vision, IET*, vol. 9, no. 3, pp. 331–339, 2015.

[4] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*. IEEE, 2007, pp. 1–8.

[5] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, Aug. 1988.

[6] M. Douze, A. Ramisa, and C. Schmid, "Combining attributes and Fisher vectors for efficient image retrieval," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2011, CVPR '11, pp. 745–752, IEEE Computer Society.

[7] M. Faraki, M.T. Harandi, and F. Porikli, "More about VLAD: A leap from Euclidean to Riemannian manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015*, June 2015, pp. 4951–4960.

[8] T. Joachims, *Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings*, chapter Text categorization with Support Vector Machines: Learning with many relevant features, pp. 137–142, Springer Berlin Heidelberg, 1998.

[9] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[10] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE Conference on Computer Vision & Pattern Recognition*, jun 2010.

[11] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *NeuroComputing*, vol. 112, pp. 172–178, 2013.

[12] S. Said, L. Bombrun, and Y. Berthoumieu, "Texture classification using Rao's distance on the space of covariance matrices," in *Geometric Science of Information (GSI)*, 2015.

[13] I. Ilea, L. Bombrun, C. Germain, R. Terebes, and M. Borda, "Statistical hypothesis test for robust classification on the space of covariance matrices," in *IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 271–275.

[14] G. Garcia and J. M. Oller, "What does intrinsic mean in statistical estimation?," *Statistics and Operations Research Transactions*, vol. 30, no. 2, pp. 125–170, 2006.

[15] K. Mader and G. Reese, "Using covariance matrices as feature descriptors for vehicle detection from a fixed camera," *ArXiv e-prints*, Feb. 2012.

[16] J. Robinson, "Covariance matrix estimation for appearance-based face image processing," *Proceedings of the British Machine Vision Conference 2005*, pp. 389–398, 2005.

[17] M. Faraki, M. T. Harandi, A. Wiliem, and B. C. Lovell, "Fisher tensors for classifying human epithelial cells," *Pattern Recognition*, vol. 47, no. 7, pp. 2348 – 2359, 2014.

[18] S. Said, L. Bombrun, and Y. Berthoumieu, "Riemannian Gaussian distributions on the space of symmetric positive definite matrices," *available on arxiv via http://arxiv.org/abs/1507.01760*, 2015.

[19] M. Moakher, "On the averaging of symmetric positive-definite tensors," *Journal of Elasticity*, vol. 82, no. 3, pp. 273–296, 2006.

[20] B. Afsari, "Riemannian lp center of mass: existence, uniqueness and convexity," *Proceedings of the American Mathematical Society*, vol. 139, no. 2, pp. 655–673, 2011.

[21] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Computer Vision ECCV 2006*, Ale Leonardis, Horst Bischof, and Axel Pinz, Eds., vol. 3952 of *Lecture Notes in Computer Science*, pp. 589–600. Springer Berlin Heidelberg, 2006.

[22] "Vision Texture Database," MIT Vision and Modeling Group. Available: http://vismod.media.mit.edu/pub/VisTex.

[23] "Outex Texture Database," Center for Machine Vision Research of the University of Oulu. Available: http://www.outex.oulu.fi/index.php?page=classification.

[24] F. Perronnin, J. Snchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Computer Vision ECCV 2010*, Kostas Daniilidis, Petros Maragos, and Nikos Paragios, Eds., vol. 6314 of *Lecture Notes in Computer Science*, pp. 143–156. Springer Berlin Heidelberg, 2010.

[25] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, "Large-scale image retrieval with compressed Fisher vectors," in *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, 2010, pp. 3384–3391.

[26] H. Hajri, I. Ilea, S. Said, L. Bombrun, and Y. Berthoumieu, "Riemannian Laplace distribution on the space of symmetric positive definite matrices," *Entropy*, vol. 18, no. 3, pp. 98, 2016.

[27] I. Ilea, L. Bombrun, C. Germain, and Y. Berthoumieu, "Texture image classification with Riemannian Fisher vectors issued from a Laplacian model," in *submitted to IEEE Image Video and Multidimensional Signal Processing (IVMSP) workshop*, 2016.