



HAL
open science

Co-occurrence Matrix of Covariance Matrices: A Novel Coding Model for the Classification of Texture Images

Ioana Ilea, Lionel Bombrun, Salem Said, Yannick Berthoumieu

► **To cite this version:**

Ioana Ilea, Lionel Bombrun, Salem Said, Yannick Berthoumieu. Co-occurrence Matrix of Covariance Matrices: A Novel Coding Model for the Classification of Texture Images. 3rd conference on Geometric Science of Information, Nov 2017, Paris, France. hal-01629171

HAL Id: hal-01629171

<https://hal.science/hal-01629171v1>

Submitted on 6 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Co-occurrence matrix of covariance matrices: a novel coding model for the classification of texture images

Ioana Ilea^{1,2}, Lionel Bombrun¹, Salem Said¹ and Yannick Berthoumieu¹

¹: Université de Bordeaux, Laboratoire IMS, Groupe Signal et Image.
{ioana.ilea, lionel.bombrun, salem.said, yannick.berthoumieu}@ims-bordeaux.fr

²: Technical University of Cluj-Napoca. *ioana.ilea@com.utcluj.ro*

Abstract. This paper introduces a novel local model for the classification of covariance matrices: the co-occurrence matrix of covariance matrices. Contrary to state-of-the-art models (BoRW, R-VLAD and RFV), this local model exploits the spatial distribution of the patches. Starting from the generative mixture model of Riemannian Gaussian distributions, we introduce this local model. An experiment on texture image classification is then conducted on the VisTex and Outex_TC000_13 databases to evaluate its potential.

Keywords: Co-occurrence matrix, Riemannian Gaussian distributions, classification, covariance matrix.

1 Introduction

Material image classification from texture contents is to assign one or more category labels to an image. It is one of the most fundamental problems in a wide range of applications such as industrial inspection [1], image retrieval [2], medical imaging [3, 4], remote sensing [5, 6], object recognition, and facial recognition [7–9]. In the general framework of image classification, feature coding techniques for bag-of-features methodologies have proven their efficiency in the recent literature. From a given feature space, bag-of-features techniques consist of first generating a codebook composed by a finite set of codewords, also called dictionary, followed by a coding step which associate to each image an activation map.

In the context of texture analysis, recent works [10–14] proposed compact and discriminative representations from localized structured descriptors in the form of region covariances, i.e. symmetric positive definite (SPD) matrices or local covariance matrices (LCM). Considering the intrinsic Riemannian geometry properties of the SPD matrix space, this paper aims at providing a competitive study of different coding techniques based on LCM codewords for texture classification.

The paper is structured as follows. Section 2 introduces the general workflow for the classification based on local descriptors. Then, sections 3 and 4 focuses

on two of its main steps, namely the dictionary learning and the coding steps. Finally, section 5 presents an experiment on texture images databases to evaluate the potential of the proposed coding model.

2 General workflow

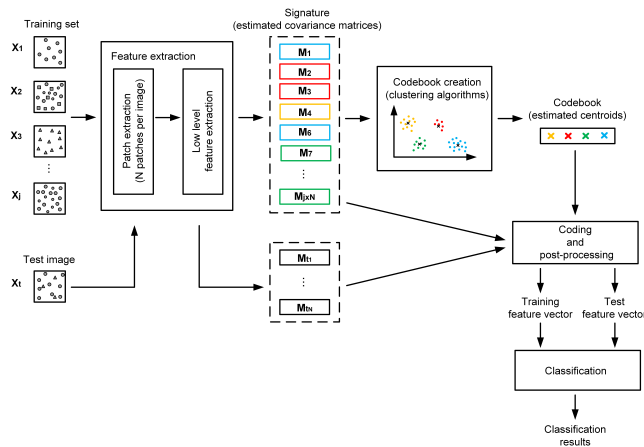


Fig. 1. Classification workflow for local features based methods.

Figure 1 presents the general workflow for the classification methods based on local features.

1. During the first step (called feature extraction), some low level features are computed from each element in the database. These descriptors are often computed on patches and as a result, a set of feature vectors (or signature) is obtained for each element in the database. These features can be covariance matrices characterizing for example the color or spatial dependencies.
2. The second step consists in the codebook creation. For that, a clustering algorithm such as the k-means or expectation maximization (EM) one is applied on the training set. By using these algorithms, the set is partitioned into a predefined number of clusters, each of them being described by parameters, such as the cluster's centroid, the dispersion and the associated weight. These estimated parameters are called codewords and are grouped in a codebook.
3. The third step is the coding stage. During this step, each signature set is projected onto the codebook space. For that, various approaches have been proposed in the literature for features being covariance matrices such as the bag of Riemannian words model (BoRW) [12], the Riemannian vectors of locally aggregated descriptors (R-VLAD) [13] and the Riemannian Fisher

vectors (RFV) [14]. Inspired by the concept of gray-level co-occurrences matrices (GLCM), the main contribution of the paper is to propose a novel coding approach which exploits the spatial arrangement between the extracted covariance matrices.

4. After the coding step, a post-processing step is classically applied, consisting in two possible normalizations, namely the ℓ_2 [15] and power normalizations [16]. These post-processing are respectively used to minimize the influence of the background information on the image signature and to correct the independence assumption made on the patches.
5. For the final classification stage, the test image is labeled to the class of the most similar training observation. In practice, classifiers such as k -nearest neighbors, support vector machine or random forest are generally employed.

The next two sections focus on the second and third step of this general workflow.

3 Dictionary learning

Let $\mathcal{M} = \{\mathbf{M}_n\}_{n=1:N}$, with $\mathbf{M}_n \in \mathcal{P}_m$, be a sample of N i.i.d observations modeled as a mixture of K Riemannian Gaussian distributions. Under the independence assumption, the probability density function (pdf) of \mathcal{M} is given by:

$$p(\mathcal{M}|\theta) = \prod_{n=1}^N p(\mathbf{M}_n|\theta) = \prod_{n=1}^N \sum_{k=1}^K \varpi_k p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \sigma_k), \quad (1)$$

where $p(\mathbf{M}_n|\bar{\mathbf{M}}_k, \sigma_k)$ is the Riemannian Gaussian density (RGD) defined on the manifold \mathcal{P}_m of $m \times m$ real, symmetric and positive definite matrices [17]. The pdf of the RGD, with respect to the Riemannian volume element, has been introduced in [17] as:

$$p(\mathbf{M}|\bar{\mathbf{M}}, \sigma) = \frac{1}{Z(\sigma)} \exp \left\{ - \frac{d^2(\mathbf{M}, \bar{\mathbf{M}})}{2\sigma^2} \right\}, \quad (2)$$

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

The codebook is hence composed by the K codewords which are the distribution parameters of each component in the mixture model defined in (1), *i.e.* the mixture weight ϖ_k , the centroid $\bar{\mathbf{M}}_k$ and the dispersion parameter σ_k . In practice, the parameters of the mixture model are estimated by considering an intrinsic k -means algorithm or an EM algorithm. For more information on the implementation of the EM algorithm, the interested reader is referred to [18].

In the experimental part, in order to ensure that each class is represented by a set of codewords, a within-class strategy is adopted to estimate the codebook. This means that a mixture model is learned for each class in the training set,

and the final codebook is obtained by concatenating each codewords (from all the classes).

Once the codebook is created, a coding step is used to encode each image in the database. For that, different strategies can be adopted such as the bag of Riemannian words (BoRW) [12], the Riemannian vectors of locally aggregated descriptors (R-VLAD) [13], the Riemannian Fisher vectors (RFV) [14] and the Co-occurrences of covariances (CoC). The next section describes each of these strategies.

4 Coding step

Let $\mathcal{M} = \{\mathbf{M}_n\}_{n=1:N}$, with $\mathbf{M}_n \in \mathcal{P}_m$, be a sample of N i.i.d covariance matrices. The aim of the coding step is to project this set \mathcal{M} onto the codebook elements.

4.1 Bag of Riemannian words (BoRW)

The bag of words (BoW) models is probably one of the most conventional methods used to encode an image. This approach is used in a wide variety of applications in computer vision and signal and image processing. But, when features are living in a non-Euclidean space such as the Riemannian manifold \mathcal{P}_m of $m \times m$ covariance matrices, this model should be readapted. For that the so-called bag of Riemannian words (BoRW) [12] and log-Euclidean bag of words (LE-BoW) [19] models have been introduced.

In these models, the data space is partitioned in K Voronoï regions by maximizing the corresponding pdf. Then, each observation \mathbf{M}_n is assigned to the cluster k , $k = 1, \dots, K$ according to:

$$\arg \max_k \varpi_k p(\mathbf{M}_n | \bar{\mathbf{M}}_k, \sigma_k), \quad (3)$$

where $p(\mathbf{M}_n | \bar{\mathbf{M}}_k, \sigma_k)$ is the RGD pdf given in (2). In practice, the homoscedasticity assumption is generally considered (*i.e.* $\sigma_k = \sigma \forall k \in [1, K]$) and the codewords are assumed to be equiprobable (*i.e.* $\varpi_k = 1/K$). Further on, for each image in the dataset, its signature is determined by computing the histogram of the number of occurrences of each codeword.

The BoRW model is a simple but effective method. Nevertheless, it suffer from a major drawback, it only counts the number of local descriptors assigned to each Voronoï region. In order to increase the classification performances, some authors have proposed some models which include second order statistics. This is the case for the R-VLAD and RFV models which are presented next.

4.2 Riemannian vectors of locally aggregated descriptors (R-VLAD)

The Riemannian version of the VLAD descriptors, called Riemannian Vectors of Locally Aggregated Descriptors (R-VLAD), has been developed in [13]. For each

cluster c_k , $k \in [1, K]$, a vector containing the differences between the cluster's centroid $\bar{\mathbf{M}}_k$ and each element \mathbf{M}_i in that cluster is computed. Next, the sum of differences concerning each cluster c_k is determined:

$$\mathbf{v}_k = \sum_{\mathbf{M}_i \in c_k} \text{Log}_{\bar{\mathbf{M}}_k} \mathbf{M}_i, \quad (4)$$

where $\text{Log}(\cdot)$ is the Riemannian logarithm mapping [20]. This model assumes two hypotheses:

- an hard assignment scheme, this means that each observation \mathbf{M}_i belongs only to one cluster c_k .
- the homoscedasticity assumption, that is $\sigma_k = \sigma, \forall k = 1, \dots, K$

In order to relax these two assumptions, the Riemannian Fisher vectors model has been introduced in [14].

4.3 Riemannian Fisher vectors (RFV)

Starting from the generative model introduced in (1), the RFV model is obtained by computing the derivative of the log-likelihood of the mixture model with respect to the distribution parameters [14].

$$\frac{\partial \log p(\mathcal{M}|\theta)}{\partial \bar{\mathbf{M}}_k} = \sum_{n=1}^N \gamma_k(\mathbf{M}_n) \sigma_k^{-2} \text{Log}_{\bar{\mathbf{M}}_k}(\mathbf{M}_n), \quad (5)$$

$$\frac{\partial \log p(\mathcal{M}|\theta)}{\partial \sigma_k} = \sum_{n=1}^N \gamma_k(\mathbf{M}_n) \left\{ -\frac{Z'(\sigma_k)}{Z(\sigma_k)} + \frac{d^2(\mathbf{M}_n, \bar{\mathbf{M}}_k)}{\sigma_k^3} \right\}, \quad (6)$$

$$\frac{\partial \log p(\mathcal{M}|\theta)}{\partial \alpha_k} = \sum_{n=1}^N [\gamma_k(\mathbf{M}_n) - \varpi_k], \quad (7)$$

where $Z'(\sigma)$ is the derivative of the normalizing factor $Z(\sigma)$ with respect to the dispersion parameter σ . The term $\gamma_k(\mathbf{M}_n)$ corresponds to the contribution of each observation \mathbf{M}_n to the cluster c_k , it is defined by:

$$\gamma_k(\mathbf{M}_n) = \frac{\varpi_k p(\mathbf{M}_n | \bar{\mathbf{M}}_k, \sigma_k)}{\sum_{j=1}^K \varpi_j p(\mathbf{M}_n | \bar{\mathbf{M}}_j, \sigma_j)}, \quad (8)$$

Note that the following parametrization of the weights in the mixture model is used in order to ensure the positivity and sum to one constraints of the weights

$$\varpi_k = \frac{\exp(\alpha_k)}{\sum_{j=1}^K \exp(\alpha_j)}. \quad (9)$$

As explained in [14], R-VLAD features are a particular case of RFV features. They are retrieved from the RFV when only the derivative with respect to $\bar{\mathbf{M}}_k$ is considered (5) and when the two hypotheses recalled in section 4.2 are assumed.

4.4 Co-occurrences of covariances (CoC)

These three models (BoRW, R-VLAD and RFV) have shown promising results, but all of these methods do not exploit one main characteristic: the spatial distribution of the patches. Inspired by the concept of GLCM to texture analysis, we introduce a novel coding approach: the co-occurrences of covariances (CoC).

For a dictionary of K codewords, the $K \times K$ co-occurrence matrix of covariance matrices for an image I describes the spatial interactions between the covariance matrices \mathbf{M}_n computed on patches separated from a distance (Δ_x, Δ_y) . The element $\mathbf{C}_{\Delta_x, \Delta_y}(k, l)$ of this co-occurrence matrix contains the number of times a covariance matrix which belongs to the codeword l occurs in the neighborhood $\mathcal{N}_{\Delta_x, \Delta_y}$ a covariance matrix which belongs to the codeword k :

$$\mathbf{C}_{\Delta_x, \Delta_y}(k, l) = \sum_{\mathbf{M}_n \in \mathcal{M}} \sum_{\mathbf{M}_p \in \mathcal{N}_{\Delta_x, \Delta_y}(\mathbf{M}_n)} \begin{cases} 1 & \text{if } \mathbf{M}_n \in c_k \text{ and } \mathbf{M}_p \in c_l \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Once the co-occurrence matrices are computed, the proximity between two CoC \mathbf{C}^1 and \mathbf{C}^2 is computed as their intersection by:

$$\sum_{k=1}^K \sum_{l=1}^K \min \left(\mathbf{C}_{\Delta_x, \Delta_y}^1(k, l), \mathbf{C}_{\Delta_x, \Delta_y}^2(k, l) \right) \quad (11)$$

This similarity measure is then used in the classification procedure.

5 Application to texture image classification

In this section, we present an application to texture image classification. The aim of this part is to evaluate the potential of the four coding models presented in section 4: BoRW, R-VLAD, RFV and CoC.

For this experiment, two databases are considered: the VisTex [21] database and the Outex_TC000_13 [22] database. The VisTex database is composed by 40 texture classes. Each class is represented by a set of 64 images of size 64×64 pixels. The Outex_TC000_13 database contains 68 texture classes, where each class is represented by a set of 20 images of size 128×128 pixels. For both databases, the feature extraction and classification steps shown in Fig. 1 are similar. We consider the same protocol as the one presented in [14]. First, covariance matrices are computed on sliding patches of size 15×15 pixels. These covariance matrices describe the interaction between 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 [10]. Then, once the images are encoded with one of the four presented model (BoRW, R-VLAD, RFV or CoC), an SVM classifier with a Gaussian kernel is used for the final classification step. In practice, the dispersion parameter of this kernel is optimized by using a cross validation procedure on the training set.

Table 1 presents the classification results in term of overall accuracy obtained on the VisTex and Outex_TC000_13 databases for the four coding models.

For the RFV model, the contribution of each parameter (centroid, dispersion, weight) is analyzed. For example, the row “RFV : ϖ ” shows the classification accuracy when only the derivatives with respect to the weights are considered to calculate the RFV (see (7)), . . . For the CoC model, an 8-neighborhood with a displacement of two pixels between the patches is considered. As observed in Table 1, the best classification results are observed for the proposed CoC model which exploits the spatial distribution of the patches. A significant gain of about 1% is observed on both VisTex and Outex TC000_13 databases compared to other state-of-the-art coding models (BoRW, R-VLAD and RFV).

Method	VisTex	Outex TC000_13
BoRW [12]	86.87 \pm 1.56	83.86 \pm 1.41
R-VLAD [13]	87.91 \pm 0.74	83.13 \pm 1.50
RFV : ϖ [14]	89.42 \pm 0.63	84.97 \pm 0.87
RFV : σ [14]	79.32 \pm 1.38	76.75 \pm 1.48
RFV : $\bar{\mathbf{M}}$ [14]	87.77 \pm 0.84	84.20 \pm 0.65
RFV : σ, ϖ [14]	82.13 \pm 1.19	79.35 \pm 1.39
RFV : $\bar{\mathbf{M}}, \varpi$ [14]	88.73 \pm 0.89	84.57 \pm 0.54
RFV : $\bar{\mathbf{M}}, \sigma$ [14]	89.43 \pm 0.79	84.01 \pm 0.65
RFV : $\bar{\mathbf{M}}, \sigma, \varpi$ [14]	89.80 \pm 0.57	84.22 \pm 0.62
CoC	91.08 \pm 0.61	85.19 \pm 0.97

Table 1. Classification results on the VisTex and Outex databases in terms of overall accuracy.

6 Conclusion

This paper has introduced a novel local model for image classification on the manifold of covariance matrices. Based on the concept of co-occurrence matrices, this local model exploits the spatial distribution of the patches, allowing to improve the classification performances compared to standard coding models (BoRW, R-VLAD and RFV).

Further works will concern the extension of such coding model to fuzzy co-occurrence matrices [23].

References

1. Liu, C., Sharan, L., Adelson, E.H., Rosenholtz, R.: Exploring features in a bayesian framework for material recognition. In: CVPR, IEEE Computer Society (2010) 239–246
2. Hiremath, P., Pujari, J.: Content based image retrieval using color, texture and shape features. 2012 18th International Conference on Advanced Computing and Communications (ADCOM) **00** (2007) 780–784
3. de Luis-García, R., Westin, C.F., Alberola-López, C.: Gaussian mixtures on tensor fields for segmentation: applications to medical imaging. Computerized Medical Imaging and Graphics **35**(1) (2011) 16–30

4. Cirujeda, P., Cid, Y.D., Müller, H., Rubin, D.L., Aguilera, T.A., Loo, B.W., Diehn, M., Binefa, X., Depeursinge, A.: A 3-D Riesz-covariance texture model for prediction of nodule recurrence in Lung CT. *IEEE Trans. Med. Imaging* **35**(12) (2016) 2620–2630
5. Zhu, C., Yang, X.: Study of remote sensing image texture analysis and classification using wavelet. *International Journal of Remote Sensing* **19**(16) (1998) 3197–3203
6. Regniers, O., Bombrun, L., Lafon, V., Germain, C.: Supervised classification of very high resolution optical images using wavelet-based textural features. *IEEE Transactions on Geoscience and Remote Sensing* **54**(6) (June 2016) 3722–3735
7. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
8. Vu, N.S., Dee, H.M., Caplier, A.: Face recognition using the POEM descriptor. *Pattern Recogn.* **45**(7) (July 2012) 2478–2488
9. Nguyen, T.P., Vu, N., Manzanera, A.: Statistical binary patterns for rotational invariant texture classification. *Neurocomputing* **173** (2016) 1565–1577
10. Tuzel, O., Porikli, F., Meer, P. In: Region covariance: a fast descriptor for detection and classification. Volume 3952 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2006) 589–600
11. Jayasumana, S., Hartley, R.I., Salzmann, M., Li, H., Harandi, M.T.: Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In: *IEEE CVPR.* (2013) 73–80
12. Faraki, M., Harandi, M.T., Wiliem, A., Lovell, B.C.: Fisher tensors for classifying human epithelial cells. *Pattern Recognition* **47**(7) (2014) 2348 – 2359
13. Faraki, M., Harandi, M.T., Porikli, F.: More about VLAD: A leap from Euclidean to Riemannian manifolds. In: *IEEE CVPR.* (June 2015) 4951–4960
14. Ilea, I., Bombrun, L., Germain, C., Terebes, R., Borda, M., Berthoumieu, Y.: Texture image classification with Riemannian Fisher vectors. In: *IEEE ICIP.* (2016) 3543 – 3547
15. Perronnin, F., Sánchez, J., Mensink, T. In: Improving the Fisher kernel for large-scale image classification. Volume 6314 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010) 143–156
16. Perronnin, F., Liu, Y., Sánchez, J., Poirier, H.: Large-scale image retrieval with compressed Fisher vectors. In: *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010.* (2010) 3384–3391
17. Said, S., Bombrun, L., Berthoumieu, Y., Manton, J.H.: Riemannian gaussian distributions on the space of symmetric positive definite matrices. *IEEE Transactions on Information Theory* **63**(4) (April 2017) 2153–2170
18. Said, S., Bombrun, L., Berthoumieu, Y.: Texture classification using Rao’s distance on the space of covariance matrices. In: *Geometric Science of Information.* (2015)
19. Faraki, M., Palhang, M., Sanderson, C.: Log-euclidean bag of words for human action recognition. *IET Computer Vision* **9**(3) (2015) 331–339
20. Higham, N.J.: *Functions of matrices: theory and computation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2008)
21. : Vision Texture Database. MIT Vision and Modeling Group. Available: <http://vismod.media.mit.edu/pub/VisTex>
22. : Outex Texture Database. Center for Machine Vision Research of the University of Oulu. Available: <http://www.outex oulu.fi/index.php?page=classification>
23. Ledoux, A., Losson, O., Macaire, L.: Texture classification with fuzzy color co-occurrence matrices. In: *IEEE ICIP.* (Sept 2015) 1429–1433