Challenge IEEE-ISBI/TCB :
Application of Covariance matrices and wavelet marginals

Florian Yger
Tokyo Institute of Technology - florian@sg.cs.titech.ac.jp

October 13, 2014

Abstract

This short memo aims at explaining our approach for the challenge IEEE-ISBI on Bone Texture Characterization. In this work, we focus on the use of covariance matrices and wavelet marginals in an SVM classifier.

1 Introduction

Texture Characterization of Bone radiograph images (TCB) is a challenge in the osteoporosis diagnosis organized for the International Society for Biomedical Imaging (ISBI) 2014. The goal of this Challenge is to identify osteoporotic cases from healthy controls on 2D bone radiograph images, using texture analysis. The dataset consists of two populations composed of 87 control subjects (CT, Figure 1 (left)) and 87 patients with osteoporotic fractures (OP, Figure 1 (right)).

As illustrated by Figure 1, textured images from the bone microarchitecture of osteoporotic and healthy subjects are very similar, making the challenge's task highly difficult .

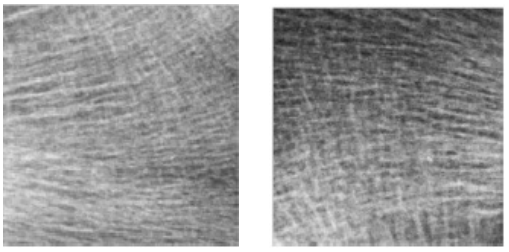

Figure 1: Example of textures of a control subject (left) and a patient with osteoporotic fractures(right).

2 Feature for textures

In our submissions to the ISBI challenge on texture classification, we have not looked for complicated application specific features or for a fancy feature selection algorithm. We rather focusd on two simple types of features, namely covariance matrices and wavelet marginals. Those submissions aimed at evaluation the features already studied to a real-life application.

2.1 Covariance matrices

Covariance matrices have been studied as image descriptor in wide variety of applications from licence plate detection [9] to

pedestrian detection [16].

For an image or a region of an image $I \in \mathbb{R}^{d_1 \times d_2}$, this approach consist in computing local features $f(x)$ (usually statistical properties) on every pixel p_{ij}. Then, for of those local descriptors, the unbiased empirical estimator of the covariance matrix is computed as :

$$C = \frac{1}{n-1} \sum_{i,j} (f_{ij} - \bar{f})(f_{ij} - \bar{f})^\top \qquad (1)$$

with $n = d_1 \times d_2$ and $\bar{f}$ being the empirical mean of f. Note that this estimator will be accurate provided that the number of samples is large enough compared to the number of features.

However, this estimator is well-known for its sensitivity to outliers. To overcome this issue, a robust estimator -Minimumum Covariance Determinant (MCD)- has been introduced in [11]. Basically, MCD aims at finding h observations (out of n) whose covariance matrix has the lowest determinant. Even if it suffered recently some controversies about its convergence properties, we use the algorithm (FastMCD [12]) that has been proposed to approximate the MCD estimator. We use the implementation provided in the LIBRA toolbox [1] for MATLAB [17]. In our experiment, when using the FastMCD algorithm, we set $\alpha = 0.9$ (meaning that the algorithm should be robust up to 10% of outliers) and $n_{\text{trial}} = 500$ the number of trial subsamples drawn from the dataset.

Concerning the local features used for computing a covariance matrix, there exists several choices. We used two variants of features used in the litterature :

- gradient based [9, 16] :

$$f_{ij} = \Big[I^{ij}, |I_x^{ij}|, |I_y^{ij}|, |I_{xx}^{ij}|, |I_{yy}^{ij}|, \ldots \\ \ldots \sqrt{(I_x^{ij})^2 + (I_y^{ij})^2}, \arctan \frac{|I_x^{ij}|}{|I_y^{ij}|} \Big]^\top \qquad (2)$$

  where I^{ij} is the intensity of the pixel (i, j) and I_x, I_{xx},.. are the intensity derivatives (first and second order along the x and y axis) and the last term is the edge orientation [2], leading to a 7×7 covariance matrix.

- Gabor based [14, 8] :

$$f_{ij} = \Big[g_1^{ij}(I), \ldots g_p^{ij}(I) \Big]^\top \qquad (3)$$

  where we have $g_1(I) = \sqrt{||\,\mathfrak{Re}(I \star g_1)||^2}$, norm of the real part of the convolution of the image I with a Gabor filter g_1. In our experiments, we used a filter bank of Gabor filters with parameters $\gamma = 1$, $\theta \in \{-\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}\}$ and $\sigma_x = \sigma_y \in \{5, 10, 20\}$, leading to a 12×12 covariance matrix.

As already noticed in the litterature [15, 9, 10, 16], covariance matrices belong to a non-Euclidean space where distances are not computed on straight lines but rather on curves lines (namely geodesics). Hence, using tools from the Riemannian geometry to analysis, it is possible to analyse those

[1] Available at `http://wis.kuleuven.be/stat/robust/LIBRA/LIBRA-home`.

[2] Note that contrary to cited paper, we did not use the pixel coordinates as it did not make sense for texture analysis and gave poor results.

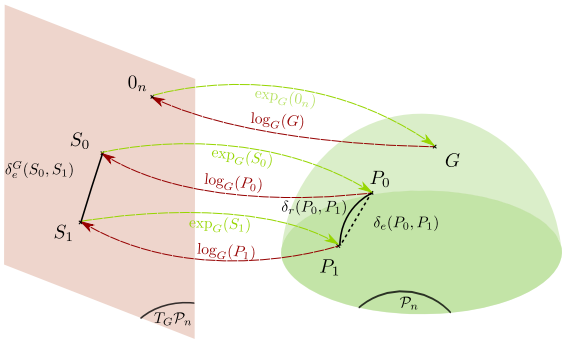

Figure 2: Mapping between the Riemannian manifold of Symmetric Definite matrices and its Tangent space at the identity. (extracted from [18])

structured data. For example, given C_i and C_j two non-singular covariance matrices, the Riemannian distance between them is :

$$\delta_R(C_i, C_i) = ||\log(C_i^{-\frac{1}{2}} C_j C_i^{-\frac{1}{2}})||_{\mathcal{F}} \quad (4)$$

with $||.||_{\mathcal{F}}$ the Frobenius norm and $\log(.)$ the matrix principal logarithm.

Recently, some authors investigated the use of such a feature for EEG signals and propose to use different kernels of the litterature to handle it [1, 18]. We intend to apply those study to covariance matrices computed on images.
In our experiments on textures (and coherently to the results in [18]), normalized LogEuclidean kernels showed the best performance in a Leave-one-out cross-validation. For two non-singular covariance matrices, this kernel $k_G(C_i, C_j)$ is defined as :

$$\frac{\mathrm{Tr}(\log(G^{-\frac{1}{2}} C_i G^{-\frac{1}{2}}) \log(G^{-\frac{1}{2}} C_j G^{-\frac{1}{2}}))}{||\log(G^{-\frac{1}{2}} C_i G^{-\frac{1}{2}})||_{\mathcal{F}} ||\log(G^{-\frac{1}{2}} C_j G^{-\frac{1}{2}})||_{\mathcal{F}}} \quad (5)$$

with the parameter G being a non-singular covariance matrix. This kernel can be understood as a normalized scalar product in a the Tangent space[3] around G (a Euclidean space where the data are mapped by the logarithm application - see Figure 2) to the space of positive definite matrices. As the property demonstrated in [18] suggest it, the choice of G induces a deformation of the shapes in the feature space. So far, two

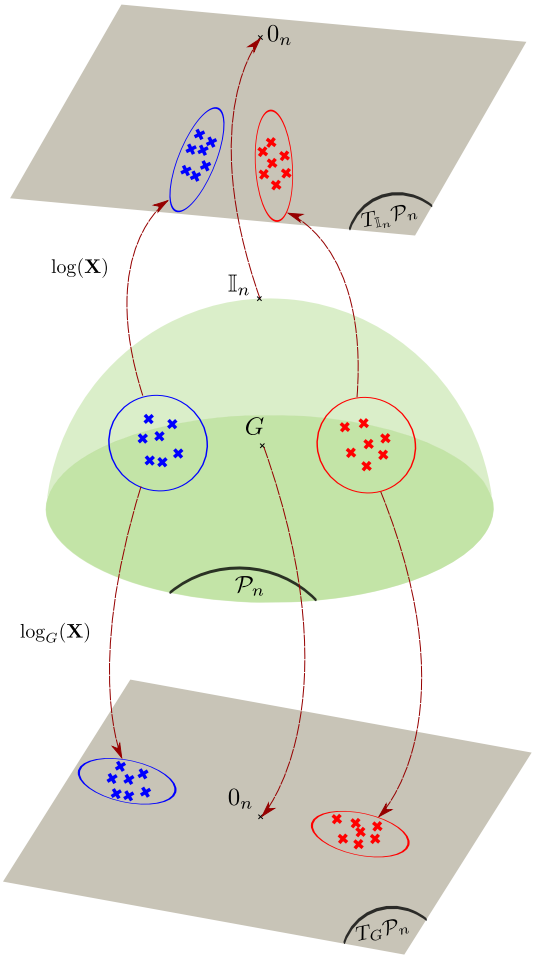

Figure 3: Illustration of the deformation induced by the choice of different reference of tangent space. (extracted from [18])

[3] It is a local linear approximation around G of the Riemannian manifold.

heuristics have been used, either the identity matrix $\mathbb{I}$ or the Riemannian mean [7] of the learning set G_{rm}.

This deformation of the geometry induced by the use of a LogEuclidean kernel is illustrated in Figure 3. Hence, it seems reasonnable to use the Riemannian mean as it may reduce the amount of distortion induced by flattening the manifold.

2.2 Wavelet marginals

Wavelet marginals are signal and image descriptors based on wavelet decomposition. This feature has been developed in order to extract frequential information for translation invariant classification of biomedical signals [3] and textures [19].

Before delving into the description of this feature, let us briefly introduce some notations in the context of one-dimensional signals[4]. Let ϕ_θ be a mother wavelet (which shape is parametrized by θ). We denote by $\theta_{\theta,s,t}$ the wavelet obtained from the mother wavelet after a dilatation at scale s and a translation t.

As originally described in [3], it is possible to extract the information contained in some frequency bands using wavelet marginals. For a signal $x \in \mathbb{R}^d$, for every $s \in [0, \cdots, \overbrace{\log_2(d)}^{J}]$, this feature is defined as :

$$m_{\theta,s}(x) = \frac{\sum_t |\langle \phi_{\theta,s,t}, x \rangle|}{\sum_t \sum_s |\langle \phi_{\theta,s,t}, x \rangle|} \qquad (6)$$

For a given signal of size d, it is possible to extract at most $\log_2(d)$ marginals. As illustrated in Figure 4, every marginal extract this information contained in a given frequency band of the signals.

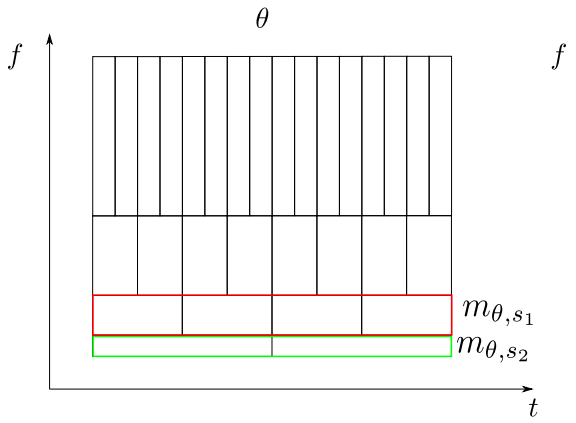

Figure 4: Illustration of the information extracted in the time-scale space by the two marginals (at scale s_1 and s_2) of a wavelet decomposition (with a waveform parametrized by θ.

Once the wavelet decomposition of a two-dimensional image is defined, it is straightforward to extend marginals to images. Let $\psi_t heta$ be a scaling function and $\phi_t heta$ its corresponding mother wavelet. For the purpose of image analysis [6], three different 2D mother wavelets are generated from the tensor product of the wavelet and the scaling function. Then, with t_x and t_y the translations along x and y axis respectively, we have the following wavelet coefficients for an image $I \in \mathbb{R}^{d \times d}$:

$$c^l_{\theta,s,t_x,t_y}(I) = \langle \phi^l_{\theta,s,t_x,t_y}, I \rangle \; l = \{1,2,3\} \quad (7)$$

As proposed in [19], by summing over the extra indices (l and t_y) in Eq.6, it is possible to extend this feature to images.

In this work, marginals are used as a baseline. Indeed, this feature is very sensitive to the waveform used for analysis

[4]For a comprehensive review of wavelet decomposition, the reader should refer to [6].

the signal. Some method has been proposed [19] in order to optimize this waveform and to select the relevant scale for classification.
For a given image, we decomposed it using a Haar wavelet[5] and then computed the marginals of the decomposition for every scale. Using the labeled dataset, we normalize the data to a zero mean and unit variance and then a linear kernel was used. Note that since unnormalized marginals are positive and sum to one, the use of χ^2 kernels[5] may be investigated.

3 Methodology

3.1 Image preprocessing

For extracting wavelet marginals, we need the images to have dyadic dimensions. Hence, we resized (using the Matlab function *imresize*) the image from 400×400 to 256×256, 128×128 or 64×64.
Finally, based on our validation results, marginals of Haar wavelets seemed to be the most efficient on 128×128 images.

On our first round of experiments, we applied gradient based covariance matrices to the raw images and obtained very low (almost random) validation results. The gradient based features being very localised, it seemed to miss some important information on the data. Hence, we resized the images by different factors in order to extract more global inforamtion. We rescaled the images to factor $\frac{1}{2}$,$\frac{1}{4}$,$\frac{1}{8}$ or $\frac{1}{16}$. On our validation procedure, a rescaling factor of $\frac{1}{8}$ (e.g. 50×50) gave the best validation performance.
However, when computing our covariance matrices, we observed very unstable results. As this estimator is not robust to outliers, we applied the FastMCD algorithm to approximate the MCD estimator. For gradient based features, we obtained a boost in the validation results (compared to the empirical covariance matrices).

For the Gabor based covariance matrices, the results have been somehow very different. We observed that applying Gabor based covariance matrices to rescaled images was giving worst validation results. This may be because the Gabor filters used were already extracting global information on the raw images. The choices of the parameters of the Gabor filters considered have been choosen based on their validation performance.
Note that contrary to the gradient based features, we did not use the FastCMD algorithm and only relied on the empirical covariance matrices of the Gabor features. Indeed, the FastCMD approximation lead to poor validation results. However, this may only indicate that we should have bette tune the parameter of the FastCMD algorithm (and raising the number trial n_{trial}.

3.2 Validation procedure

The rules of the competition specified that the competitors had to use an SVM classifier[6]. We tuned the hyperparameter C of the classifier using a Leave-one-out (LOO) cross-validation procedure. This

[5]The wavelet decomposition was performed using Wavelab 850 toolbox for Matlab available at : `http://statweb.stanford.edu/~wavelab`

[6]We used the SVM toolbox for MATLAB - available at `http://asi.insa-rouen.fr/enseignants/~arakoto/toolbox/index.html`.

feature image size	CmdMat-grad 50×50	CovMat-gab 400×400	Marginal-Haar 128×128
kernel type kernel parameter	LogEuclidean identity	LogEuclidean Riemannian mean	linear -
validation LOO accuracy	83.22% 77.59%	90.63% 74.14%	64.26% 60.34%

Table 1: Properties and mean accuracy over the Leave-one-out cross-validation procedure for our methods.

parameter could take values in the set $\{1, 10, 100, 1000, 10000, 100000\}$.

When we produced two variants of the same methods (for example, two preprocessing different for the raw images or same feature with different kernels), we selected the variant that achieved the best mean accuracy over the LOO procedure. We also report as *validation*, the mean accuracy of the learned classifiers on the training dataset.

We sum up the obtained results in Tab 1 and the properties of the proposed approach.

4 Conclusion and perspectives

Based on the validation results, it was difficult to choose one method from the three proposed. Indeed, the big gap of accuracy between the LOO and validation results led us to fear for overfitting. On the other hand, despite worst results, using the *Marginal-Haar* features shows closer validation and LOO accuracy criterion.

After having consulted the competition organizers, we submitted the three methods and obtained surprizing results.

4.1 Competition results

We first report our final results as announced by the organizers [7]

In this Tab 2, we report the criteria used by the competition organizers :

TP - True Positive : number of subjects with Osteoporosis correctly identified

FP -False Positive : number of Control subjects incorrectly identified

TN - True Negative : number of Control subjects correctly identified

FN - False Negative : number of subjects with Osteoporosis incorrectly identified

Sn - Sensitivity : defined as $Sn = \frac{TP}{TP+FN}$

Sp - Specificity : defined as $Sp = \frac{TN}{FP+TN}$

From those criteria, the challenge organizers derived six other criteria used to rank the submitted methods. In Tab 2, we report the mean rank on those 6 criteria as communicated by the challenge organizers.

From the gap between the first and blind results, it clearly appears that both covariance based methods overfitted and obtained

[7] The final results are available at `http://www.univ-orleans.fr/i3mto/results`

Method	First results						Blind results						
	TP	FP	TN	FN	Sn	Sp	TP	FP	TN	FN	Sn	Sp	rank
Marginal-Haar	36	20	38	22	0.62	0.66	19	10	19	10	0.66	0.66	1
CmdMat-grad	54	7	51	4	0.93	0.88	16	15	14	13	0.55	0.48	5
CovMat-gab	46	7	51	12	0.79	0.88	13	14	15	16	0.45	0.52	6

Table 2: Published results on the TCB competition for the three proposed methods.

deceiving results.
It should be stated that the Mariginal based method has been ranked first on every of the 6 criteria used by the organizers.

4.2 Perspectives

As stated in the introduction, we have not applied state-of-the art features in texture classification but rather tried to apply previously proposed work. Indeed, it should be noted the recently proposed scattering transform [13, 2] may be a more powerful texture descriptor than what we proposed.

It should also be noted that Wavelet marginals have been used in a rather different setting than their original proposition [19]. Indeeed, we restricted ourself to the use of a single Haar wavelet basis but since there is a strong impact on the choosen wavelet parametrizing a marginal, we should have validated carefully this choice.
In the original method, the wavelet parameter was selected through an MKL based approach that could not be applied for this competition since the rules restricted the use of SVM classifiers only (implicitly forbidding MKL methods). Moreover, as the final results suggest it, the main issue in this competition resides in overfitting, so an MKL approach having more degree of liberty, its should be very carefully tuned.

In the same line of thought, combining different features (through an MKL method) has shown very good practical results in [4]. For a real world application where enough data are available, this would be a very promising future work. Yet, in a context of data competition (with only limited data), such an MKL approach may lead to overfitting.

Acknowledgments

Part of this work has been carried in the laboratory LITIS with the support of a grant from LeMOn ANR-11-JS02-10 and it was finished in Tokyo Institute of Technology with the support of a JSPS fellowship.

References

[1] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Classification of covariance matrices using a riemannian-based kernel for bci applications. *Neurocomputing*, 112:172–178, 2013.

[2] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1872–1886, 2013.

[3] Dario Farina, Omar Feix do Nascimento, Marie-Françoise Lucas, and Christian Doncarli. Optimization of wavelets for classification of movement-related cortical potentials generated by variation of force-related parameters. *Journal of neuroscience methods*, 162(1):357–363, 2007.

[4] Peter Gehler and Sebastian Nowozin. On feature combination for multiclass object classification. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 221–228. IEEE, 2009.

[5] Bernard Haasdonk and Claus Bahlmann. Learning with distance substitution kernels. *Pattern Recognition*, pages 220–227, 2004.

[6] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[7] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.

[8] Hélio Palaio and Jorge Batista. A kernel particle filter multi-object tracking using gabor-based region covariance matrices. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 4085–4088. IEEE, 2009.

[9] Fatih Porikli and Tekin Kocak. Robust license plate detection using covariance descriptor in a neural network framework. In *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS)*, page 107, 2006.

[10] Fatih Porikli, Oncel Tuzel, and Peter Meer. Covariance tracking using model update based means on riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2006.

[11] Peter J Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984.

[12] Peter J. Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

[13] Laurent Sifre and Stéphane Mallat. Combined scattering for rotation invariant texture analysis. In *European Symposium on Artificial Neural Networks*, 2012.

[14] Jing Yi Tou, Yong Haur Tay, and Phooi Yee Lau. Gabor filters as feature images for covariance matrix on texture classification problem. In *Advances in Neuro-Information Processing*, pages 745–751. Springer, 2009.

[15] Oncel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 589–600, 2006.

[16] Oncel Tuzel, Fatih Porikli, and Peter Meer. Human detection via classification on riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007.

[17] Sabine Verboven and Mia Hubert. Libra: a matlab library for robust analysis. *Chemometrics and intelligent laboratory systems*, 75(2):127–136, 2005.

[18] Florian Yger. A review of kernels on covariance matrices for bci applications. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, pages 1–6. IEEE, 2013.

[19] Florian Yger and Alain Rakotomamonjy. Wavelet kernel learning. *Pattern Recognition*, 44(10):2614–2629, 2011.