Medical Image Classification via 2D color feature based Covariance Descriptors

Pol Cirujeda and Xavier Binefa

Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain,
{pol.cirujeda, xavier.binefa}@upf.edu

Abstract. In these notes we present an image classification method which has been submitted to the ImageCLEF 2015 Medical Classification challenge. The aim is to classify images from 30 heterogeneous classes ranging from diagnose images coming from different acquisition techniques, to various biomedical publication illustrations. The presented work is intended to be a proof of concept of how our method, which uses only visual information, performs in the modelling of such image classes. Our approach uses 1^{st} and 2^{nd} order color features obtained at a whole image level. These features are considered as samples of a multidimensional statistical distribution, and a distinctive signature of the represented image can be built in the form of a Covariance-matrix based descriptor. The Riemannian manifold structure of such descriptors can be exploited in order to formulate an image classification methodology. Despite the challenging task due to unbalanced classes and image homogeneity, the obtained results in the task place our method on the top of the most accurate ones using purely visual features. This asserts the feasibility of our methodology and proves that its performance can be on par with other methods which use also complementary textual features for complex image retrieval.

Keywords: Covariance descriptor, Medical image, classification, retrieval

1 Introduction

Medical image classification provides a challenge on the identification of similar medical images: this is an interesting problem due to the subtle changes between different image sources. For instance, inside the range of microscopy images there exist different acquisition devices (light, electron, fluorescence or transmission) which are able to capture different tissue details. Despite of that, the resemblance between image cues is high and poses a challenging problem from a classification perspective [6].

The ImageCLEF Medical Classification challenge [8] provides a benchmark to test the impact of different image classification and feature selection methods in retrieval, specially those using visual and/or textual information. We are presenting our results in the medical subfigure classification task, which provides 30

different classes including diagnose images (radiology, visible light photography, microscopy, etc.) and also generic biomedical illustrations. More details can be found in [5].

In the Computer Vision research area, many pattern recognition methods have been developed for image classification and retrieval. Most of them include the development of content and feature selection functions, or the usage of keypoint extractors and associated descriptors which can be later categorized by supervised classification methods (Support Vector Machines, Boosting, Neural Networks, etc). Our presented approach is based in our ongoing research in Covariance-based descriptors, and we are specially motivated by the demanding conditions found in the different images of the medical classification subtask. Our method provides a simplistic formulation, which provides a discriminative signature for a whole image according to the variation of different features at its pixels. We are particularly interested in seeing if this proposed description, using purely visual information, is discriminative enough.

The following sections of this report present an overview of our methodology, the results obtained on the train data and the own challenge, and a final discussion about some aspects of the presented approach and associated future work.

2 Methodology

An inspection of the provided images of this medical classification task makes evident that class separation from purely visual cues is not a trivial task. Different image sources might share visual features, or suffer from a lack of discriminative salient cues (see Fig. 1). Nevertheless, this also yields to our first intuition of what should be taken into account. First of all, there are several information cues that are equally important: not only texture patterns, but also color, sparsity, structure features... And in a second place, even more important than the features themselves: the modelling must take into account all the feature interactions together. That is, a diagram figure in a medical publication can be in grayscale just as an electron microscopy image, but structural features in a diagram contain pure lines or geometrical shapes which are not present on a biological tissue captured by the microscope. At the same time, different microscopy devices might capture similar natural tissue patterns, but for instance a visible light microscope can capture a different range of color spectra than a transmission microscope. Therefore, in an analogy with a natural visual perceptual system, our goal is to model the space of different visual cues and their joint relationships, and correlate them to the wide range of image classes.

2.1 Visual features Covariance-based descriptor

An ideal image representation must encode all images in a common compact, size invariant notation regardless of the different image sizes. The description must

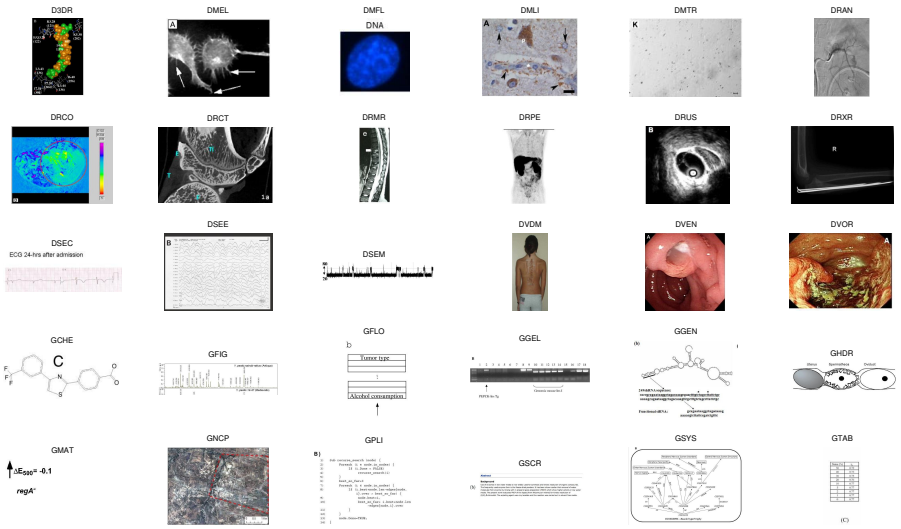

Fig. 1. Example of different samples of the different 30 classes present on the ImageCLEF Medical classification task. Please refer to [5] for more details on class hierarchy and terminology.

also be robust to intraclass spatial transformations, such as rotations, and if possible it should not depend on computationally loading intermediate stages, such as keypoint extractions. The intuition behind our method is not to use image features themselves, but rather than that observe the features along the complete image and consider them as unstructured samples of a multidimensional statistical distribution, using their covariance as a descriptive signature.

Covariance matrices were introduced as descriptors in the Computer Vision domain by Tuzel *et al.* in [7] where they presented an object recognition method for 2D color images. In our ongoing research we have extended this framework to other domains such as 3D object recognition in unstructured point clouds [3], gesture recognition in depth image sequences [2] or also tissue classification in 3D CT medical images [4]. By their construction, covariance-based descriptors are robust to noisy inputs and lose structural information about the observed features. Their representation capability is based on the statistical notion of covariance as a measure of how several random variables change together –a set of visual cues for any image in our case. Therefore, the proposed descriptor characterizes a given distribution of feature variations along the image, rather than using feature absolute values, which is independent of the number of used samples (the image size). This provides invariance to size and spatial rigid transformations such as rotations.

In order to formally define this 2D color feature based Covariance Descriptors, we denote a feature selection function $\Phi(I)$ for a given image I as:

$$\Phi(I) = \{\phi_{x,y} \, \forall x, y \in I\}, \tag{1}$$

which provides a set of feature vectors $\phi_{x,y}$ for each one of the pixel coordinates $\{x, y\}$ inside all the image I. These 11-dimensional feature vectors are expressed as:

$$\phi_{x,y} = \left[x, y, R_{x,y}, G_{x,y}, B_{x,y}, |I_x|, |I_y|, |I_{xx}|, |I_{yy}|, \sqrt{I_x^2 + I_y^2}, \arctan \frac{|I_x|}{|I_y|}\right], \quad (2)$$

and include the pixel coordinates, the different RGB color values, first and second order image intensity derivatives and their magnitude and pixel curvature. These cues provide information about the color distribution of a given image class, as well as their texture patterns and visual structure –as found in the first and second order gradient and curvature features. Then, for a given color image I the associated Covariance Descriptor can be obtained as:

$$Cov\left(\Phi(I)\right) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\phi_{x,y} - \mu\right) \left(\phi_{x,y} - \mu\right)^T, \quad (3)$$

where μ is the vector mean of the set of vectors $\{\Phi\}$ within the image I.

The resulting 11×11 matrix Cov is a symmetric matrix where the diagonal entries will represent the variance of each feature channel, and the non-diagonal elements represent their pairwise covariance, as seen in Fig. 2. This provides a signature of how feature behave in a characteristic way for each one of the images of the different classes.

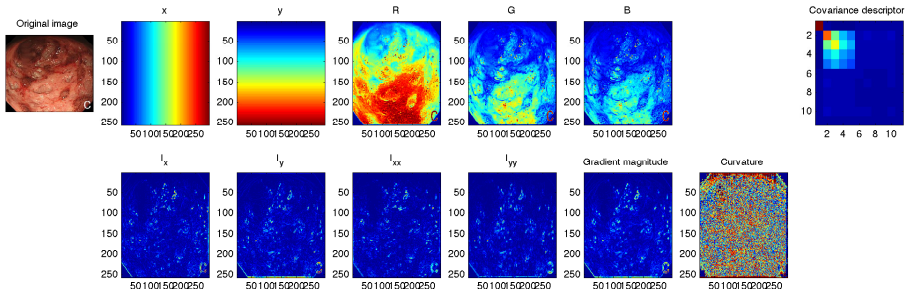

Fig. 2. Different cues involved in the descriptor building for an image of the endoscopy class (leftmost subimage). The resulting Covariance Descriptor is shown in the rightmost subfigure. Images of the same class share similar Covariance Descriptor signatures, while images from classes with different color distributions and shape features have differentiated descriptors.

2.2 Riemannian geometry of the descriptor space

Covariance Descriptors have the form of covariance matrices which, besides providing a compact and flexible representation, causes them to lie in the Riemannian manifold of symmetric definite positive matrices Sym_d^+. This has a major

impact on their interest as descriptive units, as their spatial variety is geometrically meaningful: samples of classes sharing similar feature characteristics will remain under close areas in this descriptor space. Nevertheless, it is important to bear in mind that this spatial distribution is non Euclidean and has to be treated with its particular Riemannian metric in order to perform analytic operations with the descriptors.

According to [1], the Riemannian manifold can be approximated in close neighborhoods by the Euclidean metric in its tangent space, T_Y, where the symmetric matrix Y is a reference projection point in the manifold. T_Y is formed by a vector space of $d \times d$ symmetric matrices, and the tangent mapping of a manifold element X to $x \in T_Y$ is made by the point-dependent $\log_Y$ operation:

$$x = \log_Y(X) = Y^{\frac{1}{2}} \log\left(Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}}\right) Y^{\frac{1}{2}}. \tag{4}$$

For computational simplicity in certain problems, the projection point can be established to the Identity matrix, and therefore the tangent mapping becomes:

$$\log(X) = U \log(D) U', \tag{5}$$

where U and D are the elements of the single value decomposition (SVD) of $X \in Sym_d^+$.

In an analogous manner, the exponential mapping of a point $y \in T_Y$ returns its original point representation Y in the Sym_d^+ manifold:

$$\exp(y) = U \exp(D) U', \tag{6}$$

One property of the projected symmetric matrices in the tangent space T_Y is that they contain only $d(d+1)/2$ independent coefficients, in their upper or lower triangular parts. Therefore it is possible to apply the vectorization operation in order to obtain a linear orthonormal space for the independent coefficients:

$$\hat{x} = vect(x) = (x_{1,1}, x_{1,2}, ..., x_{1,d}, x_{2,2}, x_{2,3}, ..., x_{d,d}), \tag{7}$$

where x is the mapping of $X \in Sym_d^+$ to the tangent space, resulting from Eq. (4). The obtained vector $\hat{x}$ will lie in the Euclidean space $\mathbb{R}^m$, where $m = d(d+1)/2$ –$\mathbb{R}^6$6 in the current approach.

This set of operations is useful for data visualization, feature selection, and for developing Machine Learning and classification techniques on top of the particular geometric space of the proposed Covariance Descriptors, specially taking into account the following Riemannian metric which expresses the geodesic distance between two points X_1 and X_2 on Sym_d^+:

$$\delta(X_1, X_2) = \sqrt{Trace\left(\log\left(X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}\right)^2\right)}, \tag{8}$$

or more simply $\delta(X_1, X_2) = \sqrt{\sum_{i=1}^{d} \log(\lambda_i)^2}$, where λ_i are the positive eigenvalues of $X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}$.

2.3 Classification via a Manifold-regularized sparse representation

For the classification of the proposed 2D color feature based Covariance Descriptors we propose a manifold-based sparse classification method which is part of our research as presented in previous approaches [2]. We intend to test the performance of this approach in the heterogeneous class distribution found in the ImageCLEF Medical classification task, and see if it is on par with other textual-based methods eventually presented to the challenge by other participants.

The topological layout of the proposed Covariance Descriptor yields to focus on a geometrically sensitive classification method which can exploit the Riemannian manifold distribution. Sparse representation based methods [9, 10] have shown a recent rise in the Machine Learning community in the context of face recognition. In this application, two key concepts are very relevant: sparsity and *collaborativeness*. They are related to the complexity of the model learning: not only because a complete set of learning samples is hardly available, but also because an unknown element can share characteristics from different classes. As this also the case in medical image retrieval, where images from a particular class might be scarce and the low-level visual cues provide a complex class definition, we propose a new sparse method formulation adapted to the manifold of 2D color based Covariance Descriptors.

The base intuition is that an unknown sample should be ideally represented, as accurate as possible, by using the smallest group of most similar samples from a learning set A. Then, a test sample in the form of a new vectorized Covariance descriptor $C \in \mathbb{R}^{66}$ can be expressed as a linear combination on top of the tangent space T of the Sym_d^+ manifold of the available set of training samples: $C = A\alpha$.

Let A be the whole set of n training samples, in its vectorized form according to eq. 7, from K different classes: $A = [A_1, A_2, ..., A_K] \in \mathbb{R}^{66 \times n}$, where each $A_i = \{vect(_i)\}$ is the set of vectorized Covariance descriptors which form the subset of training samples for the class i. And let $\alpha = [\alpha_1, \alpha_2, ..., \alpha_K]$ be a vector of weights corresponding to each one of the training samples in A. Then, the sparsity restriction on α can be achieved via its L2 norm minimization, proposing a manifold-aware minimization constraint which relaxes the computational expense of the method and adds numerical stability:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\{ \|C - A\alpha\|_2^2 + \|D\alpha\|_2^2 \right\} \tag{9}$$

where D is a diagonal matrix of size $n \times n$ which allows the imposition of prior knowledge on the solution with respect to the training set, using the Riemannian metric defined in eq. (8). This term contributes also on making the least squares solution stable, and on introducing forehand sparsity conditions to the vector $\hat{\alpha}$ as well. D is defined as:

$$D = \begin{pmatrix} \delta(A_1', C') & & 0 \\ & \ddots & \\ 0 & & \delta(A_n', C') \end{pmatrix} \tag{10}$$

where A_i' and C' are the unvectorized covariance descriptors for training and test samples respectively. The solution to the sparse collaborative representation, $\hat{\alpha}$, can be calculated by the following derived expression according to [10]:

$$\hat{\alpha} = \left(A^T A + D^T D\right)^{-1} A^T C \tag{11}$$

Finally, the classification label of the test sample C can be obtained by observing the regularized reconstruction residuals from the resulting sparse vector $\hat{\alpha}$:

$$class(C) = \underset{i}{\operatorname{argmin}} \left\{ \frac{\|C - A_i \hat{\alpha_i}\|_2}{\|\hat{\alpha_i}\|_2} \right\} \tag{12}$$

3 Results

The evaluation score used on the task performance assessment is the classification accuracy ratio for all the classes, computed as the ratio of true positives and negatives over the total number of samples. We collect the top results in Table 3, which are also publicly available on the challenge website [1].

Method	**Features**	**True positive ratio**
Participants 1	Visual + text	67.60
Participants 1	Only visual	60.91
Our method	Only visual	52.98
Participants 3	Only visual	45.63

Table 1. Top accuracy performances after submission evaluation of the ImageCLEF Medical Classification task. Our method accuracy is placed after the most accurate method. Using only visual features we are close to the best method, which also exploits textual information associated to the training samples.

Before the submission of the task, we tested our method on the training data set, using a 10-fold cross-validation. Each fold was adapted so at least 20% of samples of each class were kept in each subset. In classes with a very low number of samples which would cause to have some folds without class representation, some samples where duplicated. Therefore, classes with very few samples where guaranteed to be balanced and represented on the training set of our classification method. After iterating the cross-validation runs, we obtained an average accuracy of 73.24 %. As we have commented in section 2.3, the presented classifier arises as a method for expressing unknown samples as the best sparse representation regarding to a learning set. Therefore, we explain this

[1] http://www.imageclef.org/2015/medical

increase on the accuracy as a direct effect of the balancing preprocessing of those classes with very few elements.

Once the groundtruth annotations of the testing set have been made publicly available, we can analyse the different Precision and Recall values for each class as presented in Table 2, and observe if there is a particular correlation between these values and the different cardinality of each class or their visual nature.

Class	D3DR	DMEL	DMFL	DMLI	DMTR	DRAN	DRCO	DRCT	DRMR	DRPE
Class #	112	60	312	266	77	7	27	6	43	4
Precision	0.5300	0.1584	0.6629	0.6810	0.3875	0	0	0	0.1579	0
Recall	0.4732	0.2667	0.7436	0.5376	0.4026	0	0	0	0.1395	0

Class	DRUS	DRXR	DSEC	DSEE	DSEM	DVDM	DVEN	DVOR	GCHE	GFIG
Class #	0	20	0	4	1	12	4	17	8	764
Precision	0	0.0526	0	0	0	0.3333	0.1250	0.0217	0.1667	0.6600
Recall	0	0.0500	0	0	0	0.1667	0.2500	0.0588	0.5000	0.8154

Class	GFLO	GGEL	GGEN	GHDR	GMAT	GNCP	GPLI	GSCR	GSYS	GTAB
Class #	6	116	173	52	8	34	0	13	66	32
Precision	0	0.4806	0	0.0857	0	0.2143	0	0.0833	0	0.1707
Recall	0	0.5345	0	0.0577	0	0.0882	0	0.0769	0	0.2188

Table 2. Analysis of the cardinality of different classes in the testing set and their associated Precision and Recall values. These are clearly affected by the unbalanced class sets, which has a direct impact on our method due to its underlying formulation.

These results assert our hypothesis of a mandatory class balancing stage in order to boost the accuracy performance of our proposed sparse classifier.

4 Conclusions and future work

The presented approach provides two main outcomes: on one side, a Covariance-based descriptor which uses only low-level visual features and requires very low computational cost for its construction. On the other side, a classification method which takes into account the geometric properties of such representation. All together, the system provides an image retrieval method which is fast and has demonstrated to be of similar accuracy levels to other methods using complementary textual information.

Despite of that, we firmly believe that this method can be further extended in the future, in many directions. Descriptor features could be extended with a codification of medical terms associated to different image classes. Thus, visual and textual feature fusion would take place within the nature of our descriptor. On

the other side, after analysing the results and the available groundtruth annotations, we have observed a major dependency of our method on class cardinality due to its sparse representation formulation. Classes with minor representation can lead to higher classification error as a consequence of the minimization formulation of our method. Therefore, we have observed that this can be solved by incorporating a class balancing stage before the sparse regularization.

So far, the participation on the ImageCLEF Medical Classification task has provided an interesting benchmark which has contributed to test our ongoing research and identify some improvements for our methodology thanks to the particular nature of the provided testing data.

References

1. Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2):411–421, 2006.
2. Pol Cirujeda and Xavier Binefa. 4DCov: A nested covariance descriptor of spatio-temporal features for gesture recognition in depth sequences. In *International Conference on 3D vision (3DV)*, 2014.
3. Pol Cirujeda, Yashin Dicente Cid, Xavier Mateo, and Xavier Binefa. A 3d scene registration method via covariance descriptors and an evolutionary stable strategy game theory solver. *International Journal of Computer Vision*, pages 1–24, 2015.
4. Pol Cirujeda, Henning Müller, Daniel Rubin, Todd A. Aguilera, Billy W. Loo Jr., Maximilian Diehn, Xavier Binefa, and Adrien Depeursinge. 3d riesz-wavelet based covariance descriptors for texture classification of lung nodule tissue in ct. In *International Conference of the IEEE Engineering in Medicine and Biology Society (to appear)*, 2015.
5. Alba García Seco de Herrera, Henning Müller, and Stefano Bromuri. Overview of the ImageCLEF 2015 medical classification task. In *Working Notes of CLEF 2015 (Cross Language Evaluation Forum)*, CEUR Workshop Proceedings. CEUR-WS.org, September 2015.
6. Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, 2004.
7. Onzel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. *European Conference on Computer Vision*, pages 589–600, 2006.
8. Mauricio Villegas, Henning Müller, Andrew Gilbert, Luca Piras, Josiah Wang, Krystian Mikolajczyk, Alba García Seco de Herrera, Stefano Bromuri, M. Ashraful Amin, Mahmood Kazi Mohammed, Burak Acar, Suzan Uskudarli, Neda B. Marvasti, José F. Aldana, and María del Mar Roldán García. General Overview of ImageCLEF at the CLEF 2015 Labs. Lecture Notes in Computer Science. Springer International Publishing, 2015.
9. John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proc. of the IEEE*, 98(6):1031–1044, 2010.

10. D Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? In *International Conference on Computer Vision*, pages 471–478, 2011.