EXPLORING RANKLETS PERFORMANCES IN MAMMOGRAPHIC MASS CLASSIFICATION USING RECURSIVE FEATURE ELIMINATION

Matteo Masotti

University of Bologna, Department of Physics
Viale Berti–Pichat 6/2, 40127, Bologna, Italy

ABSTRACT

The ranklet transform is a recent image processing technique characterized by a multi–resolution and orientation selective approach similar to that of the wavelet transform. Yet, differently from the latter, it deals with the ranks of the pixels rather than with their gray–level intensity values. In this paper ranklets are used as classification features for a mammographic mass classification problem. Their performances are explored recursively eliminating some of the less discriminant ranklets coefficients according to the cost function of a Support Vector Machine (SVM) classifier. Experiments show good classification performances even after a significant reduction of the number of ranklet coefficients.

1. INTRODUCTION

The ranklet transform has been first introduced in [1] as an application to face detection. As described there, the ranklet coefficients produced by the application of the ranklet transform to an image are characterized by some interesting properties. First, they are non–parametric features. In fact, the ranklet transform deals with the ranks of the pixels rather than with their gray–level intensity values, namely, given $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ pixels, the value of each $\boldsymbol{p}_i$ is replaced with the value of its order among all the other pixels. Second, they are multi–resolution and orientation selective features. In fact, similarly to the bi–dimensional Haar wavelet transform, the vertical, the horizontal and the diagonal ranklet coefficients can be computed at different resolutions by means of a suitable stretch and shift of the compact supports.

In this paper ranklets are applied to a mammographic mass classification problem. Masses are thickenings of the breast tissue with size ranging from 3 mm to 30 mm often associated with the presence of a breast tumor. Similarly to what has been done in a couple of our previous works [2, 3], given a dataset composed of images representing both the mass class and the non–mass class, the ranklet transform is taken on each image, then the transformed images are classified by means of a trained Support Vector Machine (SVM). The novelty in this paper is that, in order to reduce the great amount of ranklet coefficients arising from the ranklet transform of each image, a feature reduction technique, known as Recursive Feature Elimination (RFE), is applied. Then the classification performances are explored as the ranklet coefficients are eliminated. Experimental results show that, with this method, the number of ranklet coefficients can be sensibly reduced without affecting the classification performances. Furthermore, an accurate analysis of the most discriminant ranklet coefficients gives interesting suggestions about which features are important for classification purposes.

The rest of the paper is organized as follows. In Section 2 an overview of the ranklet transform is given. Section 3 provides detailed informations about SVM and RFE. The experiments performed and the results achieved are discussed in Section 4. Conclusions are drawn in Section 5.

2. THE RANKLET TRANSFORM

Suppose that an image is constituted by $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ pixels. The ranklet transform is defined by first splitting the N pixels into two subsets T and C of size $N/2$, thus assigning half of the pixels to the subset T and half to the subset C. The two subsets T and C are defined being inspired by the three bi–dimensional Haar wavelet supports, as shown in Fig. 1. In particular, for the vertical Haar wavelet support, the two subsets T_V and C_V are defined. Similarly, for the horizontal Haar wavelet support the two subsets T_H and C_H are defined, whereas for the diagonal Haar wavelet support the two subsets T_D and C_D are defined.

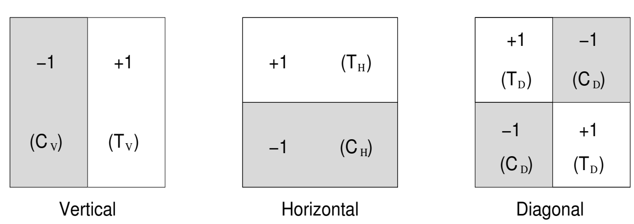

Fig. 1. The bi–dimensional Haar wavelet supports.

Submitted to *International Conference on Image Processing 2005, September 11–14, 2005, Genova, Italy*

The second step consists in computing—and then normalizing in the range $[-1,+1]$—the number of pixel pairs $(\boldsymbol{p}_m, \boldsymbol{p}_n)$, with $\boldsymbol{p}_m \in \mathsf{T}$ and $\boldsymbol{p}_n \in \mathsf{C}$, such that the intensity value of $\boldsymbol{p}_m$ is higher than the intensity value of $\boldsymbol{p}_n$. This must be done for each orientation, namely vertical, horizontal and diagonal. Notice that calculating this quantity requires approximately $O(N^2)$ operations, thus huge computational times. However, it can be demonstrated [1] that the same quantity can be calculated in approximately $O(NLogN)$ operations in the following way:

$$R_j = \frac{\sum_{\boldsymbol{p} \in \mathsf{T}_j} \pi(\boldsymbol{p}) - \frac{N}{4}(\frac{N}{2}+1)}{\frac{N^2}{8}} - 1, \quad j = V, H, D \quad (1)$$

where $\sum_{\boldsymbol{p} \in \mathsf{T}_j} \pi(\boldsymbol{p})$ is the sum of the pixel ranks in T_j, with $j = V, H, D$. The geometric interpretation of the thus obtained ranklet coefficients R_j, with $j = V, H, D$, is quite simple. Suppose that the image we are dealing with is characterized by a vertical edge, with the darker side on the left, where $\mathsf{C_V}$ is located, and the brighter side on the right, where $\mathsf{T_V}$ is located. Then R_V will be close to $+1$ as many pixels in $\mathsf{T_V}$ will have higher intensity values than the pixels in $\mathsf{C_V}$. Conversely, R_V will be close to -1 if the dark and bright side are reversed. At the same time, horizontal edges or other patterns with no global left–right variation of intensity will give a value close to 0. Analogous considerations can be drawn for the other ranklet coefficients, R_H and R_D.

The correspondence between the bi–dimensional Haar wavelet transform and the ranklet transform leads directly to the extension of the latter to its multi–resolution formulation. Similarly to what is usually done for the bi–dimensional Haar wavelet transform, the ranklet coefficients at different resolutions are computed simply stretching and shifting the bi–dimensional Haar wavelet supports. This means that the multi–resolution ranklet transform of an image is a set of triplets of vertical, horizontal and diagonal ranklet coefficients, each one corresponding to a specific resolution and shift of the bi–dimensional Haar wavelet supports.

3. CLASSIFICATION AND FEATURE SELECTION

3.1. Support Vector Machine

SVM constructs a classifier from a set of l training examples, consisting of labeled patterns $(\boldsymbol{x_i}, y_i) \in \mathbf{R}^N \times \{\pm 1\}$, $i = 1, \ldots, l$, see [4]. The classifier aims to estimate a function $f : \mathbf{R}^N \rightarrow \pm 1$, from a given class of functions, such that f will correctly classify unseen test examples $(\boldsymbol{x}, y)$. An example is assigned to the class $+1$ if $f(\boldsymbol{x}) \geq 0$ and to the class -1 otherwise.

SVM selects hyperplanes in order to separate the two classes. Among all the separating hyperplanes, SVM finds the Maximal Margin Hyperplane (MMH), namely the one that causes the largest separation among the two classes:

$$f(\boldsymbol{x}) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i (\boldsymbol{x} \cdot \boldsymbol{x}_i) + b\right) \quad (2)$$

The coefficients α_i and b are calculated by solving the following quadratic programming problem:

$$\begin{cases} \text{minimize} & J = \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) y_i y_j - \sum_{i=1}^{l} \alpha_i \\ \text{subject to} & \sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \end{cases} \quad (3)$$

where C is a regularization parameter selected by the user and J is the cost function. The classification of a pattern $\boldsymbol{x}$ is therefore achieved according to the values of $f(\boldsymbol{x})$ in (2). It is worth mentioning that the hyperplane (2) is determined only by a small fraction of training examples. These vectors, named support vectors, are those with a distance from the MMH equal to half the margin.

In the more general case in which the data are not linearly separable in the input space, a non–linear transformation $\phi(\boldsymbol{x})$ is used to map the vectors into a higher–dimensional space. The product $K(\boldsymbol{x_i}, \boldsymbol{x_j}) \equiv \phi(\boldsymbol{x_i}) \cdot \phi(\boldsymbol{x_j})$ is called kernel function. Admissible and typical kernels are:

$$\begin{cases} K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i}^T \boldsymbol{x_j} & \text{Linear} \\ K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\gamma \boldsymbol{x_i}^T \boldsymbol{x_j} + r)^d, \gamma > 0 & \text{Polynomial} \\ K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp(-\gamma \|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2), \gamma > 0 & \text{RBF} \\ K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \tanh(\gamma \boldsymbol{x_i}^T \boldsymbol{x_j} + r) & \text{Sigmoid} \end{cases} \quad (4)$$

where γ, r and d are kernel parameters selected by the user.

3.2. Recursive Feature Elimination

RFE is a general method for eliminating features which are responsible of small changes in the classifier's cost function, see [5]. In the specific case of non–linear SVM, the cost function to minimize is:

$$J = \frac{1}{2}\boldsymbol{\alpha^T H \alpha} - \boldsymbol{\alpha^T 1} \quad (5)$$

where $\boldsymbol{H}$ is the matrix with elements $y_i y_j K(\boldsymbol{x_i}, \boldsymbol{x_j})$ and $\mathbf{1}$ is an l–dimensional vector of ones. In order to compute the change in the cost function, by removing the feature f, one has to compute the matrix $\boldsymbol{H}(-f)$, where the notation $(-f)$ means that the feature f has been removed. The variation in the cost function J is thus:

$$\Delta J(f) = \frac{1}{2}\boldsymbol{\alpha^T H \alpha} - \frac{1}{2}\boldsymbol{\alpha^T H}(-f)\boldsymbol{\alpha} \quad (6)$$

The feature corresponding to the smallest $\Delta J(f)$ is then removed, SVM is trained once again with the new smaller set of features and finally tested. The procedure can thus be iterated feature after feature until a reasonable small number of features survives or the performances of the classifier start degrading.

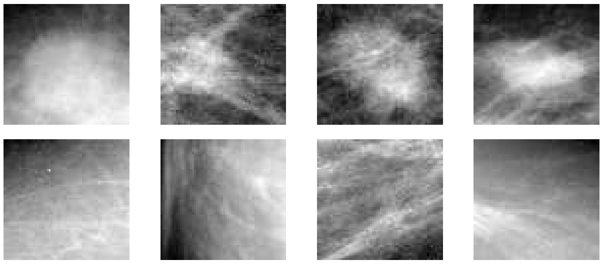

Fig. 2. The two classes of images used. Mass class (top) vs. non–mass class (down).

4. EXPERIMENTS AND RESULTS

4.1. Dataset

The images used represent both masses and non–masses. They are extracted from the mammograms of the Digital Database for Screening Mammography (DDSM), see [6], then resized to 64×64 pixels. The total number of images used amounts to 6000 and is partitioned in 1000 images representing the mass class and 5000 images representing the non–mass class. Notice that the images used in this paper are exactly those used to evaluate the classification performances in our previous works [2, 3]. In Fig. 2 some images are shown.

4.2. Ranklet Coefficients As Classification Features

The starting point of our experiments is Fig. 3, where the best classification results obtained by means of ranklet–based features are compared to those obtained by means of pixel–based and wavelet–based features. In particular, the 256 pixel–based features are obtained resizing the original 64×64 pixels images to 16×16 pixels and thus taking their gray–level intensity values. The 3000 wavelet–based features are obtained applying a redundant wavelet transform to the original 64×64 pixels images and thus taking the wavelet coefficients. Finally, the 1428 ranklet–based features are obtained resizing the original 64×64 pixels images to 16×16 pixels, applying a multi–resolution ranklet transform and thus taking the ranklet coefficients: in particular, multi–resolution is achieved by stretching the Haar wavelet supports to dimensions 16×16, 8×8, 4×4 and 2×2 pixels. For more details see [2, 3].

The performances are compared using Receiver Operating Characteristic (ROC) curves generated by moving the hyperplane of the SVM solution. This is achieved by changing the threshold b introduced in (2). Then, the fraction of true positives and false negatives for each choice of b is computed. Each single point of the ROC curves is thus obtained by averaging the results of a 10–folds cross–validation technique applied to the entire dataset.

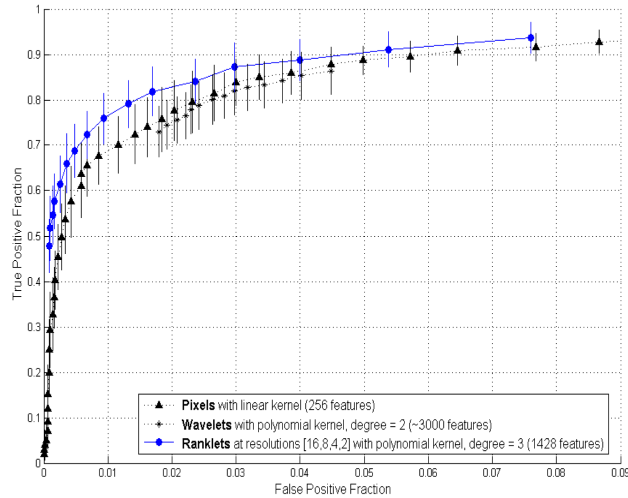

Fig. 3. Ranklets performances compared to pixel–based and wavelet–based features.

The results achieved by the three different approaches are all definitely interesting, yet the ranklet–based features achieve slightly better results. For example, focusing on a false positive fraction value of approximately 3%, the true positive fraction values are $(84 \pm 5)\%$ for the pixel–based features, $(82 \pm 5)\%$ for the wavelet–based features and finally $(87 \pm 5)\%$ for the ranklet–based features.

4.3. Ranklet Coefficients Elimination

In order to study whether it is possible to reduce the number of ranklet coefficients without sensibly affecting the classification performances, RFE is applied. The iterative procedure adopted is the following:

1. Train SVM for each fold
2. Test SVM for each fold
3. Compute the ranking criterion (6) for each feature in each fold
4. Compute a ranking list, common to all folds, by averaging the ranking position of each feature in each fold
5. Remove the feature with the smallest rank in the ranking list

It is evident from the results shown in Fig. 4 that, with this technique, the number of ranklet coefficients can be significantly reduced without affecting the classification performances. In fact, reducing the number of ranklet coefficients from 1428 down to 1000, or 200, the classification performances remain practically unaffected or, at least, they seem to take some benefit from the dimensionality reduction.

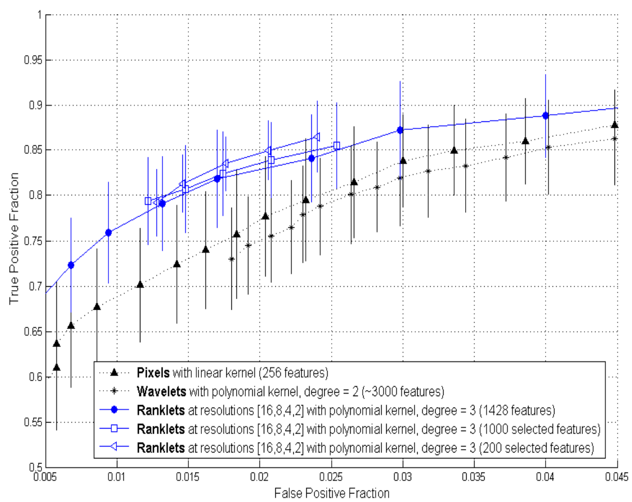

Fig. 4. Ranklets performances after the application of RFE.

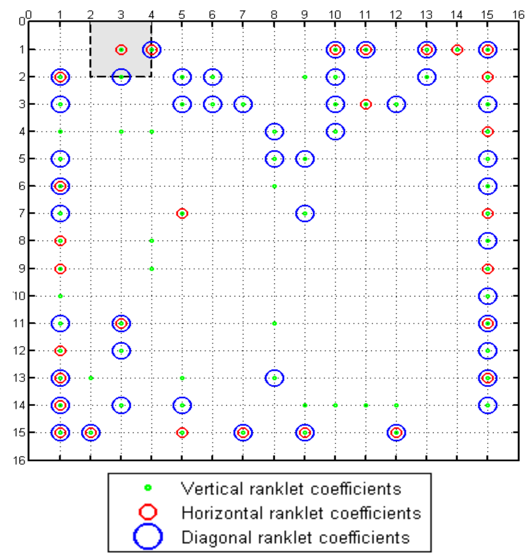

Fig. 5. Ranklet coefficients at resolution 2×2 pixels after the 500 most relevant are selected. The dashed gray square represents the Haar wavelet supports at this resolution.

Interesting considerations, about which ranklet coefficients are the most discriminating ones in this two–class classification problem, can be drawn. In Fig. 5, for example, the ranklet coefficients produced by the ranklet transform at resolution 2×2 pixels—after a selection of the 500 most relevant is made—are shown. It is evident that, at this resolution, the most discriminant ranklet coefficients are those near the borders of the image, thus those codifying the contour information of the image. That is reasonable, in fact, the main difference between the two classes at that resolution is that masses have sharp edges near the borders of the image, whereas non–masses have not. On the contrary, as the resolution decreases, it is possible to create similar plots showing that the most important ranklet coefficients are those near the center of the image, thus those codifying the symmetry information of the image, rather that its contour information. That seems to be reasonable too, since at that resolution the main difference is that masses appear approximately as symmetric circular structures centered on the image, whereas non–masses appear as less definite structures.

5. CONCLUSIONS

In this paper, a recent family of non–parametric, multi–resolution and orientation selective features, called ranklets, is applied to a mammographic mass classification problem, thus achieving good classification performances. Furthermore, by using a feature reduction technique known as RFE, the dimensionality of the classification problem is reduced from 1428 down to 200 features without affecting the classification performances.

6. REFERENCES

[1] F. Smeraldi, "Ranklets: orientation selective non-parametric features applied to face detection," in *Proceedings of the 16th International Conference on Pattern Recognition, Quebec, QC*, August 2002, vol. 3, pp. 379–382.

[2] E. Angelini, R. Campanini, E. Iampieri, N. Lanconelli, M. Masotti, and M. Roffilli, "Testing the performances of image representations for mass classification in digital mammograms," *Pattern Recognition, Submitted*, Pre–print available at http://amsacta.cib.unibo.it/archive/00000859.

[3] M. Masotti, "A ranklet–based image representation for mass classification in digital mammograms," *Pattern Recognition, Submitted*, Pre–print available at http://amsacta.cib.unibo.it/archive/00000860.

[4] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, 1995.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[6] M. Heath, K. W. Bowyer, D. Copans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," *Digital Mammography: IWDW2000 5th International Workshop on Digital Mammography*, pp. 212–218, 2000.