EXPLORING RANKLETS PERFORMANCES IN MAMMOGRAPHIC MASS CLASSIFICATION USING RECURSIVE FEATURE ELIMINATION

Matteo Masotti

University of Bologna, Department of Physics, and INFN–Bologna
Viale Berti–Pichat 6/2, 40127, Bologna, Italy
E–mail: matteo.masotti@bo.infn.it
Web: http://www.bo.infn.it/∼masotti/

ABSTRACT

The ranklet transform is a recently developed 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 pixels' ranks rather than with their gray–level intensity values. In this work, the ranklet coefficients resulting from the application of the ranklet transform to regions of interest (ROIs) found on breast radiographic images are used as classification features to determine whether ROIs contain mass or normal tissue. Performances are explored recursively eliminating some of the less discriminant ranklet coefficients according to the cost function of a support vector machine (SVM) classifier. Experiments show good classification performances (A_z values of 0.976 ± 0.003) even after a significant reduction of the number of ranklet coefficients.

1. INTRODUCTION

In 2002, the ranklet transform was developed by Smeraldi as a specific tool for face detection problems [1]. The ranklet coefficients resulting from the application of the ranklet transform to different faces were therein used with success as input features for a distance–based classification scheme. The reasons for the excellent results achieved by this approach rely mainly in the notable properties which characterize ranklet coefficients. Ranklet coefficients are, for instance, non–parametric. The ranklet transform, in fact, deals with pixels' ranks rather than with their gray–level intensity values; i.e., supposing that $(p_1, \ldots, p_N)$ pixels are given, the intensity value of each p_i is replaced with the value of its order among all the other pixels. Secondly, ranklet coefficients are multi–resolution and orientation–selective. Similarly to the bi–dimensional Haar wavelet transform [2], in fact, ranklet coefficients can be calculated at different resolutions and orientations (i.e., vertical, horizontal, and diagonal) by means of a suitable stretch and shift of the oriented compact supports used for their computation.

In this work, the ranklet transform is used to determine whether regions of interest (ROIs) found on breast radiographic images contain mass or normal tissue. Masses are thickenings of the breast tissue with size ranging from 3 mm to 30 mm often associated with the presence of breast cancer, whereas normal tissue represents healthy tissue. In one of our recent works, a database of ROIs representing both the mass class and the non–mass class (i.e., normal tissue) was collected; the ranklet transform was then applied to each ROI and the transformed ROIs discriminated by means of a previously trained support vector machine (SVM) classifier [3]. Experiments demonstrated that ranklet features perform better than pixel and wavelet features evaluated in one of our previous works on the same database [4]. With the intention of reducing the great amount of ranklet coefficients arising from the ranklet transform of each ROI, a feature reduction technique known as SVM recursive feature elimination (SVM–RFE) is applied herein. Classification performances are then explored as the ranklet coefficients are eliminated. Experimental results show that, by following this approach, the number of ranklet coefficients can be sensibly reduced without affecting classification performances. Furthermore, an accurate analysis of the most discriminant ranklet coefficients gives interesting insights about which features are important for classification purposes.

2. DATA AND METHODS

2.1. ROI database

The ROIs used in this work are extracted from radiographic images of the publicly available digital database for screening mammography (DDSM) collected by the University of South Florida [5]. In Fig. 1 some of them are shown. The total number of ROIs analyzed amounts to 6000. A partition of 1000 ROIs representing diagnosed mass tissue is collected by extracting square crops centered on the location of each annotated mass from the DDSM benign and malig-

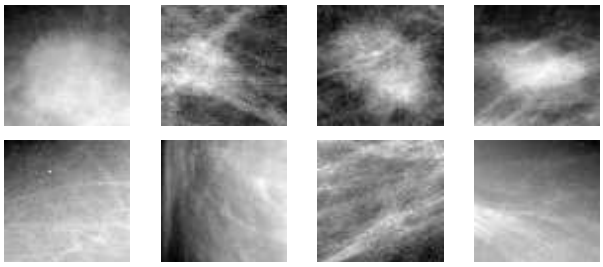

Fig. 1. ROIs belonging to different classes. Mass class (top) vs. non–mass class (bottom).

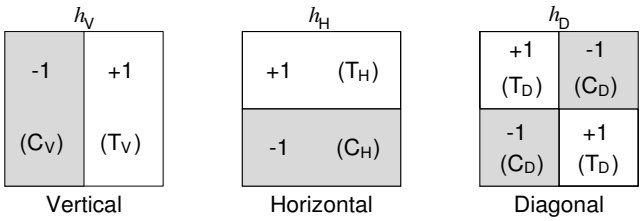

Fig. 2. The three Haar wavelet supports h_{V}, h_{H}, and h_{D}. From left to right, the vertical, horizontal, and diagonal Haar wavelet supports.

nant cases. For the non–mass class, a partition of 5000 ROIs representing normal tissue is collected by randomly extracting square crops from the DDSM normal cases. Since SVM deals exclusively with dimensionally homogeneous feature vectors, all ROIs are resized to an arbitrarily prefixed size. After initial tests, this size is chosen to be 64×64 pixels.

2.2. The ranklet transform

Suppose that an image is constituted by $(p_1, \ldots, p_N)$ 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 Haar wavelet supports shown in Fig. 2. For the vertical Haar wavelet support, the two subsets $\mathsf{T_V}$ and $\mathsf{C_V}$ are defined. Similarly, for the horizontal Haar wavelet support the two subsets $\mathsf{T_H}$ and $\mathsf{C_H}$ are defined, whereas for the diagonal Haar wavelet support the two subsets $\mathsf{T_D}$ and $\mathsf{C_D}$ are defined. The definition of the aforementioned Haar wavelet supports forms the basis for the orientation–selective property of the ranklet transform.

The second step consists in computing and normalizing in the range $[-1, +1]$ the number of pixel pairs (p_m, p_n), with $p_m \in \mathsf{T}$ and $p_n \in \mathsf{C}$, such that the intensity value of p_m is higher than the intensity value of p_n. This is done for each orientation, namely vertical, horizontal, and diagonal. Calculating this quantity requires $O(N^2)$ operations, thus huge computational times. Nevertheless, it can be demonstrated that the same quantity can be calculated in $O(NLogN)$ operations as follows [1]:

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

where $\sum_{p \in \mathsf{T}_j} \pi(p)$ is the sum of the pixels' ranks $\pi(p)$ in T_j. The geometric interpretation of the derived ranklet coefficients R_j 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}. In this context, the use of the pixels' ranks, rather than their intensities, forms the basis for the non–parametric property of the ranklet transform.

Finally, the close correspondence between the Haar wavelet transform and the ranklet transform leads directly to the extension of the latter to its multi–resolution formulation. Similarly to what is done for the bi–dimensional Haar wavelet transform, the ranklet coefficients can be computed at different resolutions by simply stretching and shifting the Haar wavelet supports. The multi–resolution ranklet transform of an image is thus a set of triplets of vertical, horizontal, and diagonal ranklet coefficients, each one corresponding to a specific stretch and shift of the Haar wavelet supports. Suppose, for example, that the horizontal and vertical shifts of the Haar wavelet supports along the horizontal and vertical dimensions of the image are of 1 pixel. The number n_{T} of triplets $R_{\mathsf{V,H,D}}$ at each resolution is thus computed as $n_{\mathrm{T}} = (I + 1 - S)^2$, where I and S represent the linear dimension of the image and that of the Haar wavelet support, respectively. Here, the possibility of computing ranklet coefficients at different resolutions forms the basis for the multi–resolution property of the ranklet transform.

2.3. SVM classifier

SVM constructs a binary classifier from a set of l training samples consisting of labeled patterns $(\boldsymbol{x_i}, y_i) \in \mathbf{R}^N \times \{\pm 1\}$, $i = 1, \ldots, l$ [6]. Taking values $+1$ or -1, the label y_i indicates the class membership (mass or non–mass) of the correspondent feature vector $\boldsymbol{x}_i$ which contains a certain number of ranklet coefficients. The classifier aims at estimating a function $f : \mathbf{R}^N \rightarrow \pm 1$, from a given class of functions, such that f will correctly classify unseen test samples $(\boldsymbol{x}, y)$. An unseen sample 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, namely the one that causes the largest separation between itself and the borderline training samples of the two classes:

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

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

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

where C is a regularization parameter and J is the cost function to minimize.

In the more general case in which data are not linearly separable in the input space, a non–linear transformation $\phi(\boldsymbol{x})$ is used to map feature vectors into a higher–dimensional space where they are linearly separable. With this approach, classification problems which appear quite complex in the original feature space can be afforded by using simple decision functions in the mapped feature space, for instance linear hyperplanes. In order to implement this mapping, the dot products $\boldsymbol{x} \cdot \boldsymbol{x}_i$ are substituted by $\phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}_i) \equiv K(\boldsymbol{x}, \boldsymbol{x}_i)$, commonly referred to as kernel function. Admissible and typical kernels are the linear kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$, the polynomial kernel $K(\boldsymbol{x}, \boldsymbol{y}) = (\gamma \boldsymbol{x} \cdot \boldsymbol{y} + r)^d$, etc.

2.4. SVM recursive feature elimination

SVM–RFE is a general method for eliminating features responsible of small changes in the classifier's cost function [7]. In the specific case of non–linear SVM, the cost function to minimize is that discussed in Eq. (3) or, more compactly:

$$J = \frac{1}{2} \boldsymbol{\alpha}^{\boldsymbol{T}} \boldsymbol{H} \boldsymbol{\alpha} - \boldsymbol{\alpha}^{\boldsymbol{T}} \mathbf{1} \tag{4}$$

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:

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

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.

2.5. Performance evaluation

In order to evaluate classification performances, a 10–folds cross–validation procedure is adopted in this study [8]. The ROI database is partitioned into 10 distinct and homogeneous folds, then SVM is trained with the collection of the first 9 folds (i.e., 900 mass crops plus 4500 non–mass crops) and tested on the fold left out (i.e., 100 mass crops plus 500 non–mass crops). Training and test are repeated 10 times by changing the test fold in a round–robin manner. Performances are analyzed using the receiver operating characteristic (ROC) methodology [9]. For each test partition, the ROC curve of the system and its associated area A_z are evaluated. An average A_z value is then obtained by averaging the 10 aforementioned A_z values. ROC curves and their associated areas A_z are estimated using the `ROCKIT` software by Metz.

3. TESTS AND RESULTS

3.1. Ranklet features

The starting point for our tests is represented by the best classification result previously reached using ranklet features and its relationship with the best classification result reached by pixel and wavelet features [3, 4]. To this purpose, in Fig. 3, the best ROC curve achieved using ranklet features (**RankletS3**$_{\{16,8,4,2\}}$) is compared to the best ones achieved using pixel (**PixHRS**) and wavelet (**OwtS2**) features. Pixel features are 256 and are obtained by (a) resizing the original 64×64 pixels ROIs to 16×16 pixels and (b) considering their gray–level intensity values. Wavelet features are 2955 and are obtained by (a) applying a redundant wavelet transform [10] to the original 64×64 pixels ROIs and (b) considering the wavelet coefficients. In particular, previous experiments indicated that higher–resolution ROIs are fundamental for wavelet features in order to reach good classification performances; that is why ROIs are not resized before the wavelet transform is performed [4]. Finally, ranklet features are 1428 and are obtained by (a) resizing the original 64×64 pixels ROIs to 16×16 pixels, (b) applying the multi–resolution ranklet transform, and (c) considering 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. As they offer the best results [3, 4], a linear SVM kernel is used in combination with pixel features, whereas polynomial SVM kernels with degree 2 and 3 are used in combination with wavelet and ranklet features, respectively.

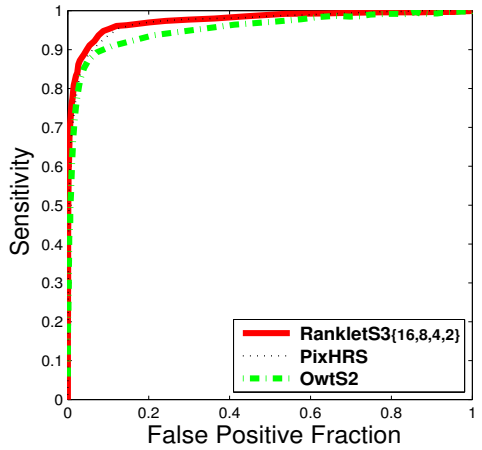

Fig. 3. ROC curves comparison. **RankletS3**$_{\{16,8,4,2\}}$ (A_z value of 0.978 ± 0.003) represents the best ranklet features previously evaluated [3]. **PixHRS** (0.973 ± 0.002) and **OwtS2** (0.956 ± 0.003) represent the best pixel and wavelet features previously evaluated [4].

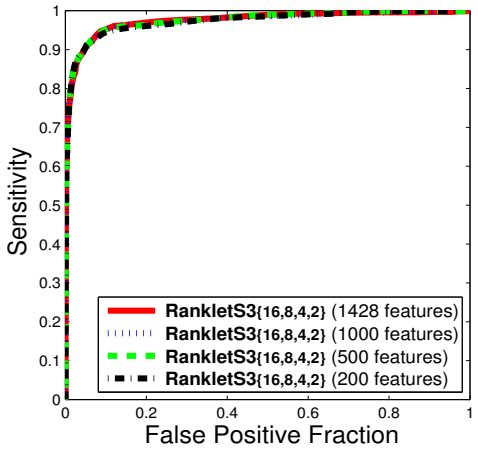

Fig. 4. ROC curves for ranklet features after the application of SVM–RFE. The differences in the A_z values achieved by the original 1428 ranklet features and the reduced ones do not achieve statistical significance.

The results reached by the three different approaches are all definitely interesting, yet ranklet features achieve slightly better results. More specifically, ranklet features achieve average A_z values of 0.978 ± 0.003, pixel features of 0.973 ± 0.002, and wavelet features of 0.956 ± 0.003. The improvement on the A_z value of ranklet features over that achieved by pixel features does not achieve statistical significance, whereas its improvement over that achieved by wavelet features is statistically relevant with two–tailed p–value < 0.0001.

3.2. Reduced ranklet features

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

1. Train SVM for each of the 10 cross–validation repetitions
2. Test SVM by performing ROC analysis for each of the 10 cross–validation repetitions and compute an average A_z value by averaging the 10 A_z values
3. Compute the ranking criterion (Eq. 5) for each feature in each of the 10 cross–validation repetitions
4. Compute an average ranking list, common to all of the 10 cross–validation repetitions, by averaging the ranking position of each feature over each of the 10 cross–validation repetitions
5. Remove the feature with the smallest rank in the average ranking list
6. Repeat until a reasonable small number of features survives or performances (i.e., average A_z value) start degrading

Two aspects of this approach deserve some deep and careful consideration. For instance, SVM must be re–trained after each feature elimination. This is reasonable, since the weight of a feature characterized by medium–low importance may be promoted by removing a correlated feature. Secondly, each cross–validation repetition is characterized by a different training set. After each training phase, thus, the computation of the ranking criterion can lead to a ranking list different for each repetition. This specifically means that the feature having smallest ranking can be different for each repetition. In order to eliminate the same feature from all the training sets, it is thus necessary to compute a ranking list common to all repetitions.

In Fig. 4, the ROC curves achieved reducing the number of ranklet coefficients by means of SVM–RFE from 1428 down to 1000, 500, and 200 are shown. A polynomial SVM kernel with degree 3 is used. As evident, the number of ranklet coefficients can be sensibly reduced without significantly affecting classification performances. The reduced 1000 ranklet features, in fact, achieve average A_z values of 0.978 ± 0.003, the reduced 500 ranklet features achieve average A_z values of 0.977 ± 0.002, and the reduced 200 ranklet features achieve average A_z values of 0.976 ± 0.003. The difference among these average A_z values and that achieved by the original 1428 ranklet features does not achieve statistical significance. On the other hand,

a smaller number of features affects more or less sensibly the computational times required by SVM for calculating the dot products in Eq. (2) and therefore assigning a crop to the mass or non–mass class. On a dual Intel Xeon 2.6 GHz PC, the original 1428 ranklet features take approximately 30 seconds for the analysis of an entire radiographic image, whereas the reduced 1000, 500, and 200 ranklet features take approximately 20, 10, and 4 seconds, respectively.

4. DISCUSSION

Some interesting considerations can be drawn about which ranklet coefficients are the most discriminating ones in this two–class classification problem. To this purpose, it is necessary to look carefully at the ranklet coefficients which survive after SVM–RFE. In Fig. 5, for example, the 200 most discriminating ranklet coefficients obtained by applying to the entire ROI database the ranklet transform at resolution 16×16, 8×8, 4×4, 2×2 pixels and then SVM–RFE are shown. Small circles represent vertical ranklet coefficients, whereas medium and large circles represent horizontal and diagonal ranklet coefficients, respectively. The dashed square represents the dimensions of the Haar wavelet supports. By looking carefully at the ranklet coefficients calculated at resolutions 2×2 and 4×4, it is evident that the majority of those surviving are located near the borders of the ROI, thus are those codifying the contour information of the ROI. This is reasonable, as the main difference between the two classes at fine resolutions is given by the presence or absence in the ROI of a boundary delimiting the bright centered nucleus of the mass. On the contrary, as the resolution decreases to 8×8 and 16×16, the most important ranklet coefficients are those near the center of the ROI, thus those codifying the symmetry information of the ROI, rather that its contour information. That is reasonable too, as at coarse resolutions the main difference is that masses appear approximately as symmetric circular structures centered on the ROI, whereas normal tissue has a less definite structure.

5. CONCLUSIONS

In this paper, a recently developed family of non–parametric, multi–resolution, and orientation selective features, called ranklets, is applied to a mammographic mass classification problem, in order to determine whether ROIs found on breast radiographic images contain mass or normal tissue. By using a feature reduction technique known as SVM–RFE, the dimensionality of the classification problem is reduced from 1428 down to 200 features without significantly affecting the classification performances (A_z values of 0.978 ± 0.003 and 0.976 ± 0.003, respectively). Furthermore, it is discussed that the true majority of the ranklet coefficients survived at the feature reduction are those produced by the ranklet transform at fine resolutions near the borders of the ROI and at coarse resolutions near the center of the ROI.

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*, 2002, vol. 3, pp. 379–382.

[2] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[3] M. Masotti, "A ranklet–based image representation for mass classification in digital mammograms," Tech. Rep. 930, University of Bologna, Department of Physics, March 2005.

[4] E. Angelini, R. Campanini, E. Iampieri, N. Lanconelli, M. Masotti, and M. Roffilli, "Testing the performances of different image representations for mass classification in digital mammograms," *International Journal of Modern Physics C*, vol. 17, no. 1, pp. 113–131, 2006.

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

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

[7] 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.

[8] R. Kohavi, "A study of cross–validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1145.

[9] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, no. 9, pp. 720–733, 1986.

[10] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi–scale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

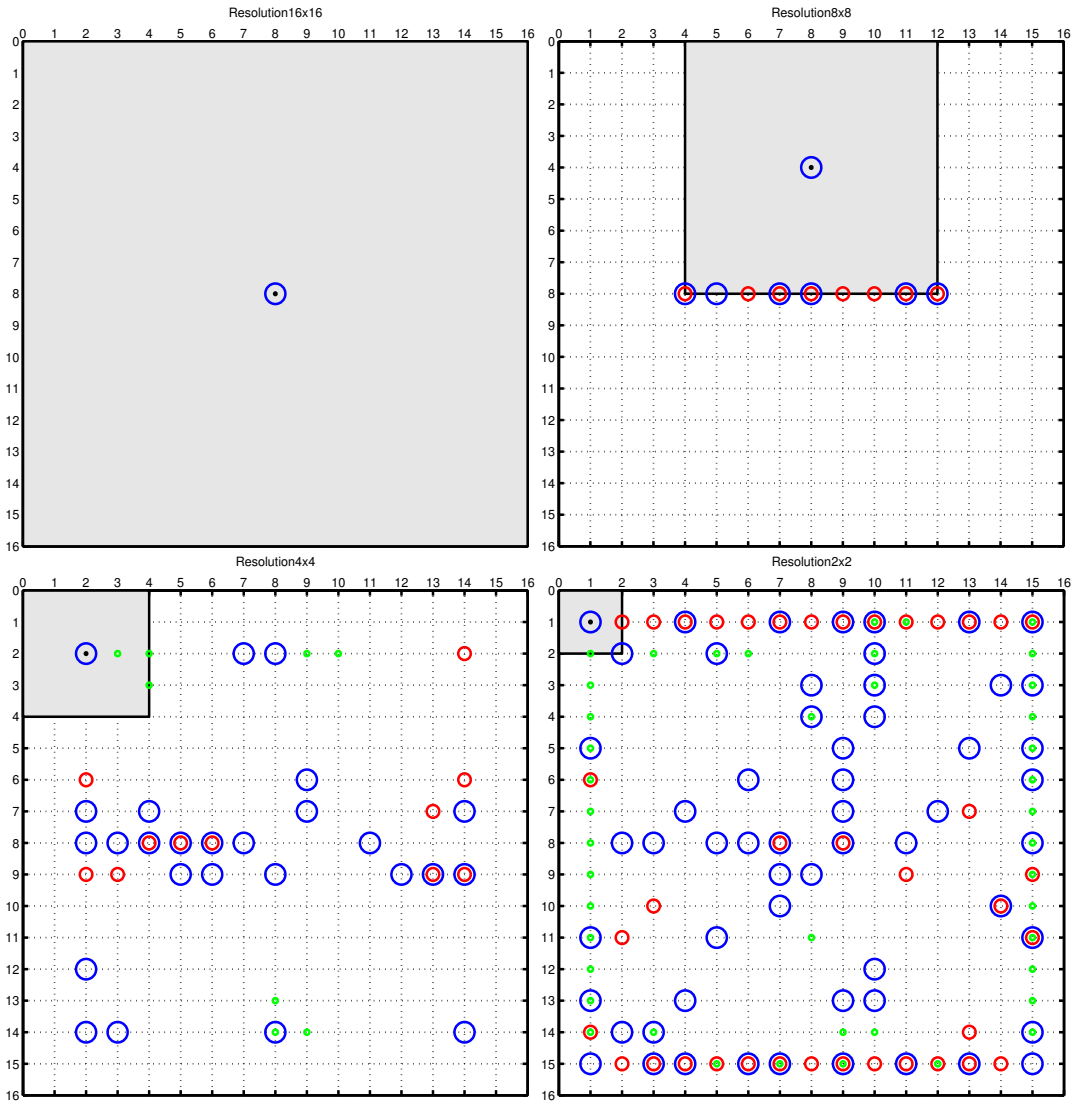

Fig. 5. Ranklet coefficients after SVM–RFE has selected the 200 most relevant ones. Small circles represent vertical ranklet coefficients, medium circles represent horizontal ranklet coefficients, large circles represent diagonal ranklet coefficients. The dashed square represents the dimensions of the Haar wavelet supports. Resolutions 16×16 (upper–left), 8×8 (upper–right), 4×4 (lower–left), and 2×2 (lower–right) are represented.