Testing the performance of image representations for mass classification in digital mammograms

Enrico Angelini‡, Renato Campanini†, Emiro Iampieri†,

Nico Lanconelli†, Matteo Masotti† *, Matteo Roffilli‡,

†Department of Physics, University of Bologna

‡Department of Computer Science, University of Bologna

Abstract

In this paper a two–class classification problem is faced. One class is constituted by tumoral masses, breast tumors with size ranging from 3 mm to 30 mm. The other class is constituted by non–masses. A Support Vector Machine (SVM) is used as a classifier. Both, masses and non–masses, are extracted from the University of South Florida (USF) mammographic image database and are presented to the classifier as crops with pixel size 64×64. In order to find the optimal solution to this problem, different featureless crops representations are evaluated. In particular, a pixel–based representation, a Discrete Wavelet Transform (DWT) representation and an Overcomplete Wavelet Transform (OWT) representation are tested.

Key words: Wavelets, Support Vector Machine, Computer Aided Detection

* Corresponding Author, *E-mail: masotti@bo.infn.it, Phone: +39 051 2095136, Fax: +39 051 2095047, Address: Viale Berti–Pichat 6/2, 40127, Bologna, Italy*

1 Introduction

Breast cancer is one of the most common causes of death among women from all over the world. An early detection of this disease is really important to increase the probability of surviving. However, due to the subtle nature of the mammographic lesions, poor radiographic image quality and eye fatigue, this task could prove really difficult for the radiologists. In order to facilitate it, Computer Aided Detection (CAD) systems have been introduced in the last years. The idea behind these systems is to automatically detect the regions suspected to be tumors. In this way, it is possible to turn the radiologists' attention directly to the suspected regions, thus facilitating the diagnosis. In Fig. 1, the CAD's mark individuating a mass is shown.

Masses are the most common lesions associated with the presence of breast tumor. They are thickenings of the breast tissue that, in the radiographic image, appear as lesions with size ranging from 3 mm to 30 mm. In order to detect them, the entire image is scanned with a resizeable window at different scales. Each sub–image scanned by the window—also known as *crop*—is then resized to an image with pixel size 64×64. Finally, each resized crop is classified as belonging to the mass class or to the non–mass class by a previously trained learning machine, namely a Support Vector Machine (SVM). For more information concerning the whole scanning scheme and the application of SVM in CAD systems for mammography, see respectively [1,2].

In this work, the whole attention is devoted to the crops classification, rather than to the scanning of the mammographic image and to the crops resizing. A two–class classification problem is faced, where the two classes, masses versus non–masses, are both constituted by crops with pixel size 64×64. In order to

find the optimal solution to this two–class classification problem, some pixel–based and wavelet–based image representations are evaluated. In the evaluation of the pixel–based image representation, the raw pixel values of each crop are used to train and to test SVM. In the evaluation of the wavelet–based image representations—both in the Discrete Wavelet Transform (DWT) and in the Overcomplete Wavelet Transform (OWT) approach—the wavelet coefficients are presented to the classifier. For this reason, the approach adopted could be considered as a featureless approach, since the training and the test of the classifier are performed without extracting any feature from the crops, thus without assuming any a priori knowledge about them.

The rest of the paper is organized as follows. In Section 2 an overview of the dataset is given. Section 3 discusses the image representations evaluated. The classification techniques adopted are discussed in Section 4. In Section 5 the methods and results for each image representation are shown. The paper is concluded in Section 6.

2 Dataset

The dataset used to evaluate both the pixel–based and the wavelet–based image representations is composed of 6000 crops with pixel size 64×64 representing the two classes, masses and non–masses, as shown in Fig. 2. The crops representing the mass class are 1000, whereas the crops representing the non–mass class are 5000. All the crops are extracted—and then resized to 64×64—from the mammographic images belonging to the Digital Database for Screening Mammography (DDSM), collected by the University of South Florida (USF), see [3]. The DDSM images are digitized with Lumisys scan-

ner at 50 μm and Howtek scanner at 43.5 μm pixel size. They have a 12–bit gray–level resolution.

3 Image representations

As mentioned in Section 1, in order to find the optimal solution to the two–class classification problem, three image representations are tested: a pixel–based, a DWT–based and an OWT–based. In the pixel–based image representation, the raw pixel values are presented to the classifier. In the DWT–based, the Haar wavelet coefficients of the 1st level are presented to the classifier. In the OWT–based, the redundant Haar wavelet coefficients of the 4th and 6th levels are presented to the classifier. In addition to these tests, the effects on the image representations performance of some pre–processing techniques, such as histogram equalization and crops resizing are evaluated. In the following, an overview of these representations and pre–processing techniques is given.

3.1 Pixel–based image representation

In the pixel–based image representation the raw pixel values are used. See Fig. 3 for an example of the pixel–based image representation.

3.2 Wavelet–based image representation

In the image processing community, the wavelet transform is a well–known technique allowing the multi–resolution analysis of images. It offers a suitable

image representation for highlighting structural, geometrical and directional features of the objects within the image, see [4].

The classical wavelet transform—also known as Discrete Wavelet Transform (DWT)—is an orthogonal transform that, through a cascade of low–pass and high–pass filters, transforms an image with pixel size $N \times N$ into $N \times N$ wavelet coefficients [5]. Each pair of filters corresponds to a decomposition level, in other words to a particular resolution of the analysis. The wavelet coefficients are divided into *approximation* coefficients, representing the image structural information, and *horizontal, vertical, diagonal* coefficients, representing respectively the horizontal, vertical and diagonal information of the image.

In order to split the information of the image on a higher number of wavelet coefficients, the redundant wavelet transform—also known as Overcomplete Wavelet Transform (OWT)—has been introduced [6]. It provides a redundant encoding of the image information through a spatially superposed wavelet analysis. For example, given an image with pixel size 64×64, the OWT produces—up to the 6th decomposition level—approximately 14 000 wavelet coefficients, whereas the application of the DWT produces $64 \times 64 = 4096$ wavelet coefficients.

3.2.1   DWT–based image representation

In the DWT–based image representation, the Haar wavelet coefficients of the 1st level are used, for a total number of 4096 wavelet coefficients for each image. See Fig. 4 for an example of the DWT–based image representation.

3.2.2 OWT–based image representation

In the OWT–based image representation, the Haar wavelet coefficients of the 4th and 6th levels are used (except the *approximation* coefficients) for a total number of approximately 3000 wavelet coefficients for each image. See Fig. 5 for an example of the OWT–based image representation.

3.3 Crops pre–processing techniques

Some image processing techniques, such as image histogram equalization and image resizing, are quite common in the imaging community, see [7]. Image histogram equalization is often used to enhance the image contrast, image resizing to decrease the image size. In order to evaluate the combined effects of these techniques, together with the previously discussed image representations, they are applied to the crops as pre–processing techniques. In the pixel–based image representation case they are applied before the SVM classification, whereas in the wavelet–based image representation case they are applied before the wavelet transform and the following SVM classification.

3.3.1 Crops histogram equalization

The process of adjusting the image intensity values—also known as histogram equalization—involves the transformation of the image intensity values so that the histogram of the equalized image approximately matches a flat histogram. A mass crop before and after histogram equalization is shown in Fig. 6.

3.3.2 Crops resizing

The process of image resizing involves the interpolation of adjacent pixels in order to estimate an image of a different size. The most common interpolation methods are the nearest–neighbor, bilinear and bicubic. In this work, a bilinear interpolation is used. This means that each pixel of the resized image is a weighted average of pixels in the nearest 2–by–2 neighborhood. A mass crop before and after resizing is shown in Fig. 7.

4 Classification

As introduced in Section 1, in order to evaluate the image representations performance, an SVM is used as a classifier. Due to the limited number of crops to train and test SVM, a cross–validation procedure is implemented. The classification results are presented in terms of the Receiver Operating Characteristic (ROC) curve of the system. In the following, an overview of SVM theory is given, together with some background information regarding the implementation of a cross–validation procedure and the evaluation of an ROC curve for a classification system.

4.1 Support Vector Machine

SVM constructs a binary classifier from a set of l training examples, consisting of labeled patterns $(\mathbf{x_i}, y_i) \in \mathbf{R}^N \times \{\pm 1\}, i = 1, \ldots, l$, see [8,9]. 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 $(\mathbf{x}, y)$. An example is assigned to the class $+1$ if $f(x) \geq 0$ and to the class -1 otherwise.

SVM selects hyperplanes in order to separate the two classes. Among all the separating hyperplanes, SVM finds the one that causes the largest separation among the decision function values for the borderline examples of the two classes. The Maximal Margin Hyperplane (MMH) is computed as a decision surface of the form:

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

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

$$\begin{cases} \text{maximize} & \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j(\mathbf{x}_i \cdot \mathbf{x}_j)y_i y_j \\ \\ \text{with} & \sum_{i=1}^{l} \alpha_i y_i = 0 \qquad\qquad 0 \leq \alpha_i \leq C \end{cases} \tag{2}$$

C is a regularization parameter, selected by the user. The classification of a pattern $\mathbf{x}$ is therefore achieved according to the values of $f(\mathbf{x})$ in (1). It is worth mentioning that in a typical classification problem the hyperplane (1) 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 $\phi(\mathbf{x})$ is used to map the input vectors into a high–dimensional space. The product $K(\mathbf{x_i}, \mathbf{x_j}) \equiv \phi(\mathbf{x_i}) \cdot \phi(\mathbf{x_j})$ is called

kernel function. Admissible and typical kernels are:

$$
\begin{cases}
K(\mathbf{x_i}, \mathbf{x_j}) = \mathbf{x_i}^T \mathbf{x_j} & \text{Linear Kernel} \\[2em]
K(\mathbf{x_i}, \mathbf{x_j}) = (\gamma \mathbf{x_i}^T \mathbf{x_j} + r)^d, \gamma > 0 & \text{Polynomial Kernel} \\[2em]
K(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma \|\mathbf{x_i} - \mathbf{x_j}\|^2), \gamma > 0 & \text{RBF Kernel} \\[2em]
K(\mathbf{x_i}, \mathbf{x_j}) = \tanh(\gamma \mathbf{x_i}^T \mathbf{x_j} + r) & \text{Sigmoid Kernel}
\end{cases}
\tag{3}
$$

where γ, r and d are kernel parameters.

4.2   Cross–validation

Cross-validation is a common procedure used to train and test a classifier when the dimensionality of the dataset is limited [10]. Given a n–dimensional dataset D, first divide the entire dataset in f homogeneus sub–datasets, also known as *folds*, $F_1, F_2, \ldots, F_f$. Train the classifier with the collection of the first $f-1$ folds, $F_1, F_2, \ldots, F_{f-1}$, then test it on F_f, the fold left over. Permute the procedure for each F_i, $i = 1, \ldots, f-1$.

As discussed in Section 2, the dataset used in this work is composed of 1000 crops representing the mass class and 5000 crops representing the non–mass class. In order to implement a 10–folds cross–validation procedure, the dataset is divided into 10 folds, each one containing 100 mass crops and 500 non–mass crops. Thus, for each permutation of the cross-validation procedure, SVM is trained with 900 mass crops and 4500 non–mass crops, then is tested on 100 mass crops and 500 non–mass crops.

4.3   ROC

The receiver operating characteristic curve analysis is a widely used method for evaluating the performance of a classifier used to separate two classes [11]. The ROC curve is a plot of the classifier's True Positive Fraction (TPF) versus its False Positive Fraction (FPF). Here the FPF is the fraction of non–masses incorrectly classified as belonging to the mass class, whereas the TPF is the fraction of masses correctly classified as belonging to the mass class. The TPF is generally known as the system *sensitivity*, the quantity 1-FPF as the system *specificity*. The best possible prediction method would yield 100% sensitivity (all true positives are found) and 100% specificity (no negatives are found).

5   Methods and results

The pixel–based, the DWT–based and the OWT–based image representations are evaluated as stand–alone image representations and together with the combined effects of the previously discussed pre–processing techniques, namely histogram equalization and resizing. A further technique—called *scaling*—is tested in combination with the ones described above. It consists of scaling correspondent features of the train and test sets in the range $[0, 1]$. The scaling coefficients are calculated for each feature during the training phase, then are used to scale correspondent features both in the train and test set. In this work, correspondent features are correspondent pixels when evaluating the pixel–based representation and are correspondent wavelet coefficients when evaluating the wavelet–based representations. This technique is very common in the pattern classification community, since it is useful in order to avoid

that features of greater value dominate those of smaller value. Furthermore, since classification depends mainly on the inner products of feature vectors, the scaling technique is useful to avoid numerical difficulties.

In order to optimize SVM to the different representations, several polynomial kernels are tested, from degree 1 up to degree 5. The performances are compared using ROC curves generated by moving the hyperplane of the SVM solution by changing the threshold b, see Eq. 1. The fraction of true positives and false negatives for each choice of b is then computed. Each single point of the ROC curves is obtained by averaging the results of a 10–folds cross–validation technique applied to the entire dataset.

In the following, the tests performed and the results obtained are described and discussed in detail. Notice the range of the axis in the performances plots of the next figures: the FPF range is $[0, 0.18]$ while the TPF range is $[0, 1]$ in order to show the most interesting parts of the ROC curves.

5.1 Pixel–based performances

For the sake of semplicity, the two pixel–based image representations tested are referred to as *PixRS* and *PixHRS*.

- *PixRS* is a pixel–based representation in combination with resizing and scaling techniques. In particular, the original crops with pixel size 64×64 are first resized to 16×16, then scaled as described above.

- *PixHRS* is a pixel–based representation in combination with histogram equalization, resizing and scaling techniques. The original crops with pixel size 64×64 are first treated with histogram equalization, then resized to 16×16 and finally scaled.

11

Fig. 8 shows the performances of *PixRS* and *PixHRS*. A linear kernel for the SVM classifier is used in both the experiments. This is in fact the one leading to the best results, when working with this image representation. Looking at the performances, *PixHRS* seems to perform significantly better than *PixRS*. This is manly due to the fact that the histogram equalization enhances the image contrast, thus probably leading to a two–class classification problem more easily separable. It is worth mentioning that the scaling technique proves to be fundamental with this image representation since, without scaling the pixels, the training of the system does not converge.

5.2 DWT–based performances

The first four DWT–based image representations tested are referred to as *Dwt*, *DwtH*, *DwtS* and *DwtHS*.

- *Dwt* is a simple DWT–based representation. The original crops with pixel size 64×64 are decomposed by DWT up to level 1 using the Haar wavelet filters.

- *DwtH* is a DWT–based representation in combination with histogram equalization. The original crops with pixel size 64×64 are first treated with histogram equalization, then decomposed by DWT up to level 1 using the Haar wavelet filters.

- *DwtS* is a DWT–based representation in combination with scaling. The original crops with pixel size 64×64 are decomposed by DWT up to level 1 using the Haar wavelet filters, then scaled.

- *DwtHS* is a DWT–based representation in combination with histogram equalization and scaling. The original crops with pixel size 64×64 are

12

first treated with histogram equalization, then decomposed by DWT up to level 1 using the Haar wavelet filters and finally scaled.

Fig. 9 shows the performances of the DWT–based representations discussed above. In this first set of experiments a linear kernel is used. It is possible to notice that there is not a significant difference among the four representations. This probably arises from the fact that the range of the wavelet coefficients is smaller than the range of the pixel values, thus scaling has a quite limited influence. Similarly, histogram equalization has a weak impact on the performances of the system, since the DWT itself enhances the contrast of the image.

The second set of tests using the DWT–based image representation is oriented in the direction of quantifying the importance of the kernel used in order to improve the classification performances. Chosen *DwtHS*, which has slightly better performances, several polynomial kernels are tested, from degree 2 up to degree 5. These tests are referred to as *DwtHS2*, *DwtHS3*, *DwtHS4*, *DwtHS5*, where the number indicates the degree of the polynomial kernel used. The performances are shown in Fig. 10. It is evident here that, differently from the pixel–based case, in which the linear kernel is the best performing one, better performances correspond to increasing values of the polynomial kernel degree.

5.3 OWT–based performances

The two OWT–based image representations evaluated are called *Owt2* and *OwtH2*, where the number, as in the previous case, indicates the degree of the polynomial kernel used.

13

- *Owt2* is a simple OWT–based representation. The original crops with pixel size 64×64 are decomposed by OWT using the Haar wavelet filters and retaining only the wavelet coefficients corresponding to the levels 4 and 6.

- *OwtH2* is an OWT–based representation in combination with histogram equalization. The original crops with pixel size 64×64 are first treated with histogram equalization, then decomposed by OWT using the Haar wavelet filters and retaining only the wavelet coefficients corresponding to the levels 4 and 6.

Fig. 11 shows the performances of the OWT–based representations discussed above. In both the experiments a polynomial kernel with degree equal to 2 is used, since this is the one leading to the best results. In these two tests, similarly to what discussed for the DWT–based image representation, there is no evidence of an outperformance of one representation with respect to the other. The reason is probably the same mentioned in Section 5.2 for the DWT case, in other words the fact that the wavelet transform itself tipically enhances the image contrast.

5.4 Image representations comparison

The previously discussed tests show that *PixRHS* is the image representation leading to the best results among all the pixel–based image representations, *DwtHS5* among all the DWT–based and *Owt2* among all the OWT–based. In Tab. 1, the results obtained with these image representations are compared. It is clear from the results that *PixRHS* has similar performances, compared to *DwtHS5* and *Owt2*. This result is quite interesting. First, it demonstrates that one of the simplest representations leads to the same results obtained with

14

more elaborated representations. This means that it is possible to avoid time–consuming image processing techniques, such as DWT or OWT, achieving similar results. Second, it demonstrates that decreasing the dimensionality of the classification problem to 256, as it is for *PixRHS*, it is possible to obtain results similar to the ones obtained by representations characterized by a much higher dimensionality, as 4096 for *DwtHS5* or approximately 3000 for *Owt2*. This means a further saving of time, since the lower is the dimensionality of the classification problem, the lower is the time used to perform the classification.

6 Conclusions and future works

The tests performed to evaluate the different combinations of representations and pre–processing techniques showed some interesting results. First, as regards the pixel–based image representation, image histogram equalization proved to be really important in order to get better performances. Image scaling proved to be fundamental for the convergence of SVM training. Linear kernel proved to be the best performing one, with this image representation. Second, as regards the wavelet–based image representations, image histogram equalization and image scaling did not seem to be fundamental. The polynomial kernels with higher degrees proved to be the best performing ones with this image representation. Nevertheless, the main result of this work is that it is possible to obtain quite close performances with specific combinations of the discussed representations and pre–processing techniques. In particular, *PixRHS* performs similarly to *DwtHS5* and *Owt2*. *PixRHS* is a pixel–based representation in combination with histogram equalization, resizing and scaling techniques, *DwtHS5* is a DWT–based representation in combination with

15

histogram equalization and scaling, while *Owt2* is an OWT–based representation. This demonstrates that, using one of the simplest representation tested, namely *PixRHS*, it is possible to obtain results quite close to the ones obtained with representations much more sophisticated, such as *DwtHS5* and *Owt2*. This literally means having the possibility to obtain similar results saving a lot of computational time.

Obviously, when dealing with a classification problem, the final aim is to achieve the best classification results, namely the best possible sensitivity and specificity, as discussed in Section 4.3. In this work, the best possible results are achieved both by sophisticated image representations—*DwtHS5* and *Owt2*—and by a much simpler image representation—*PixRHS*—in much less time. The next step will be in the direction to understand whether there is the possibility to improve the performances. Since the pixel–based representation has probably achieved its best performances at all, the future research will be mainly concentrated on the wavelet–based representation. Two aspects will be investigated. First, a deeper study of the wavelet–based representation will be conducted. Basically, different wavelet–based multi–resolution techiques will be evaluated, such as ridgelets, curvelets and steerable filters, in combination with some wavelet–based de–noizing techniques, such as wavelets shrinking and quantizing. Second, in order to investigate whether it is possible to reduce the dimensionality of the problem—thus reducing the computational time—without affecting the classification performances, some techniques such as Principal Component Analysis (PCA), Indipendent Component Analysis (ICA) and Recursive Feature Elimination (RFE) will be evaluated.

References

[1] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, M. Roffilli, A novel featureless approach to mass detection in digital mammograms based on support vector machines, Physics in Medicine and Biology 49 (6) (2004) 961–975.

[2] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, D. Romani, An svm classifier to separate false signals from microcalcifications in digital mammograms, Physics in Medicine and Biology 46 (6) (2001) 1651–1663.

[3] M. Heath, K. W. Bowyer, D. Copans, R. Moore, P. Kegelmeyer, The digital database for screening mammography, Digital Mammography: IWDW2000 5th International Workshop on Digital Mammography (2000) 212–218.
URL http://marathon.csee.usf.edu/Mammography/Database.html

[4] E. J. Stollnitz, T. D. DeRose, D. H. Salesin, Wavelets for computer graphics: A primer, part 1, IEEE Computer Graphics and Applications 15 (3) (1995) 76–84.

[5] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.

[6] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, Shiftable multi–scale transforms, IEEE Transactions on Information Theory 38 (1992) 587–607.

[7] K. R. Castleman, Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, USA, 1996.

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

[9] V. Vapnik, Statistical Learning Theory, J. Wiley, 1998.

[10] L. Tarassenko, Guide to Neural Computing Applications, Butterwoth–Heinemann, 1998.

[11] C. E. Metz, Roc methodology in radiologic imaging, Investigative Radiology 21 (1986) 720–733.

About The Author—Enrico Angelini graduated cum laude in Computer Science at the University of Bologna in 2001. During 2002 he was involved in a project of in–vitro carcinogenesi study with cellular automata. Now he is working on advanced pattern recognition methods for mass detection in digital mammography.

About The Author—Renato Campanini graduated in Physics with a thesis in experimental particle physics. In 1976 he received the Italian National Prize "Guglielmo Marconi" for physics. Since 1977 to 1986 he has been working in data analysis in particle physics at CERN (Geneva). His interests include medical imaging and microarray data analysis.

About The Author—Emiro Iampieri received the Laurea degree in Physics from the University of Bologna in 2001. Since 2002 he is mainly involved in the software development and optimization of authomatic systems finalized to the early detection of microcalcifications and masses in mammography. His interests include image analysis and software development.

About The Author—Nico Lanconelli graduated in Physics and obtained a Ph.D. in Physics in 2003. Since 1998 he is involved in projects concerning diagnostic imaging in mammography, the development of systems for the automatic detection of suspect areas in digital images, and investigation of apparatus for the early diagnosis of breast cancer.

About The Author—Matteo Masotti received the Laurea degree cum laude in Physics from the University of Bologna in 2001. He is going to receive the Ph.D. degree in March 2005, working on the optimization of wavelet techniques for mass detection in mammography. His interests include wavelets, pattern recognition and image registration.

About The Author—Matteo Roffilli graduated in Computer Science. Since 2002 he is a Ph.D. student in Computer Science at the University of Bologna. His research interests are machine learning, artificial intelligence, medical imaging and high performance computing. In the last years he worked also on object recognition for visual surveillance and devised portable multimedia applications.

List of Figures

Table 1

Classification results comparison. TPF values obtained by the best performing image representations, for FPF values approximately equal to 0.01, 0.02, 0.03, 0.04 and 0.05, are shown.

	FPF $\sim$ 0.01	FPF $\sim$ 0.02	FPF $\sim$ 0.03	FPF $\sim$ 0.04	FPF $\sim$ 0.05
PixRHS	$.70 \pm .06$	$.77 \pm .07$	$.84 \pm .05$	$.86 \pm .05$	$.89 \pm .03$
DwtHS5	$.68 \pm .08$	$.76 \pm .05$	$.81 \pm .04$	$.85 \pm .04$	$.86 \pm .03$
Owt2	-	$.75 \pm .05$	$.82 \pm .05$	$.85 \pm .05$	$.87 \pm .05$
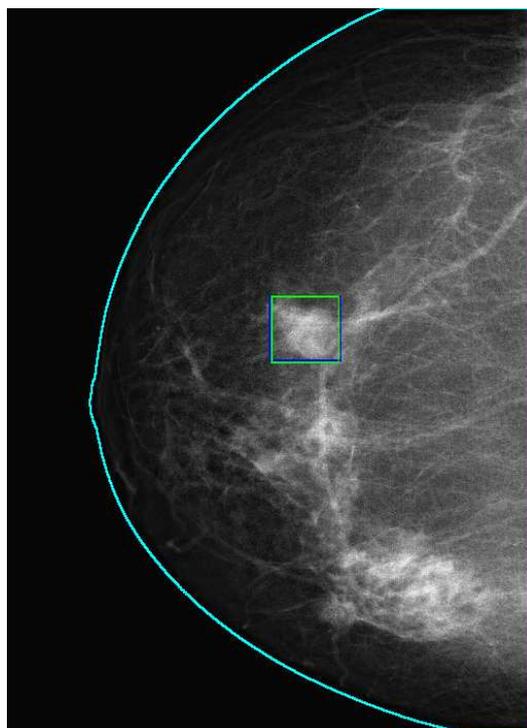
Fig. 1. Mammographic image. The square mark is the CAD's mass detection.
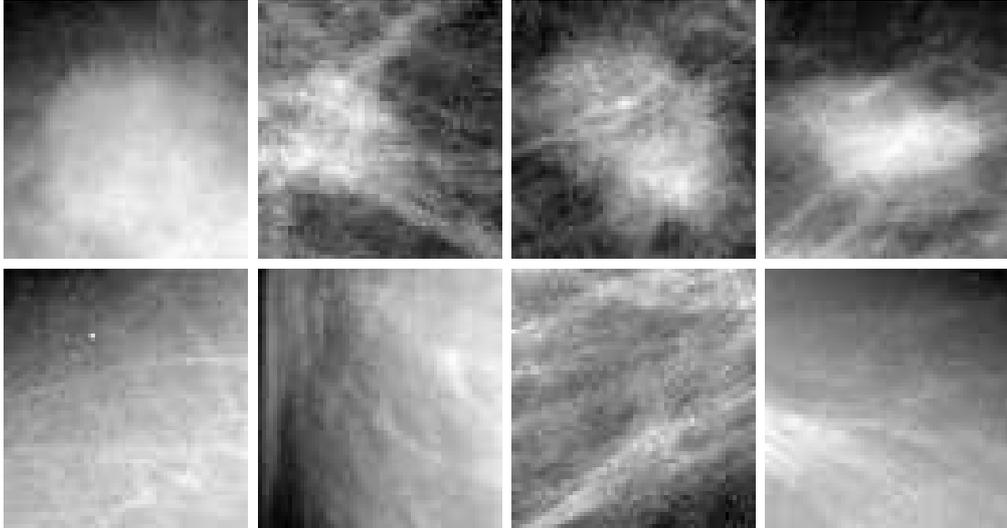
Fig. 2. The two classes. Mass class (top) vs. non–mass class (down).
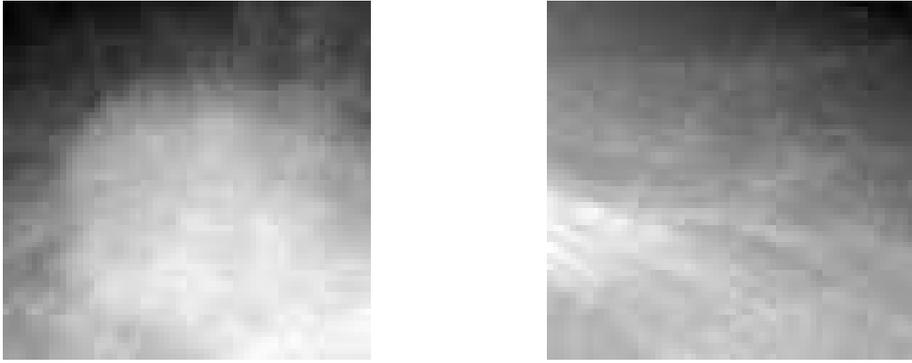
Fig. 3. Pixel–based image representation. Mass (left) vs. non–mass (right).
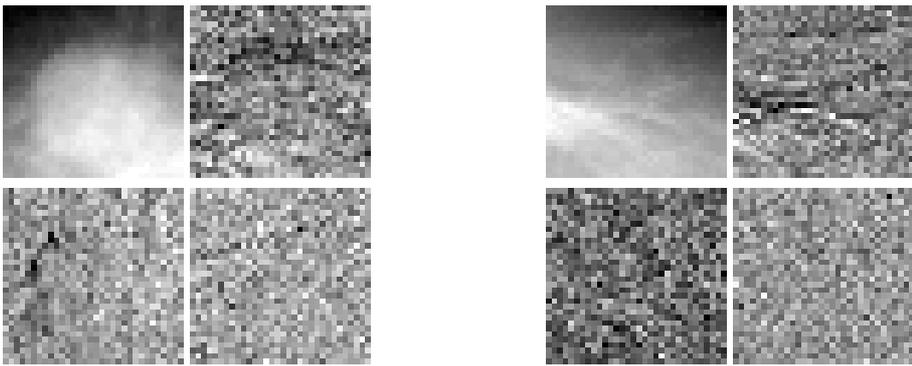
Fig. 4. DWT–based image representation. Mass (left) vs. non–mass (right). The *approximation* (upper–left), *horizontal* (upper–right), *vertical* (lower–left) and *diagonal* (lower–right) wavelet coefficients are shown both for mass and non–mass.
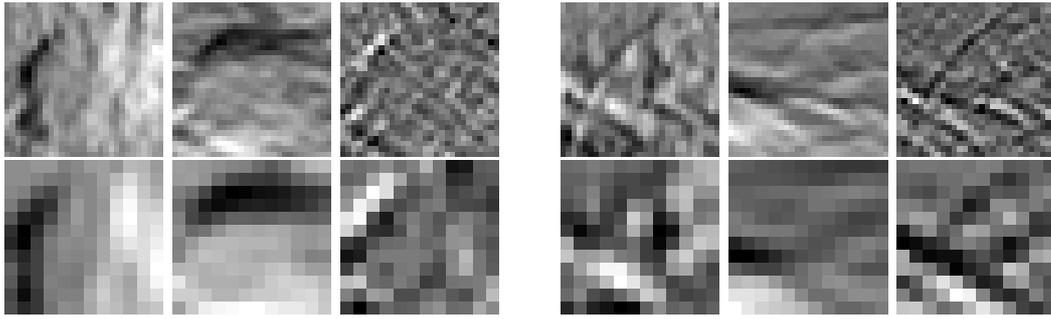
Fig. 5. OWT–based image representation. Mass (left) vs. non–mass (right). The *horizontal* (left), *vertical* (center) and *diagonal* (right) wavelet coefficients of level 4 (up) and 6 (down) are shown both for mass and non–mass.
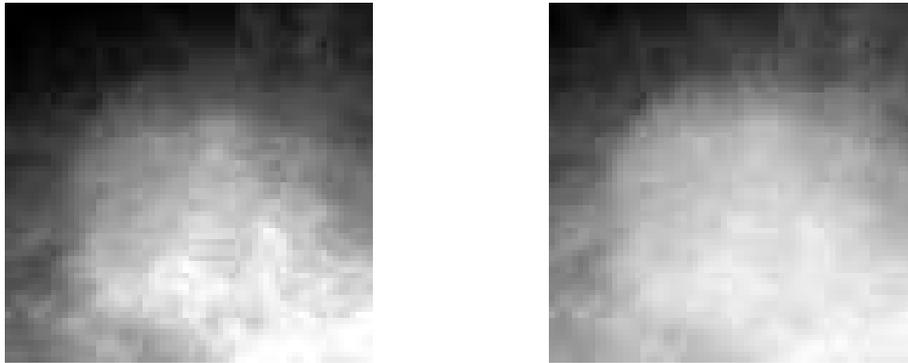
Fig. 6. Crops histogram equalization. Equalized mass (left) vs. non equalized mass (right).
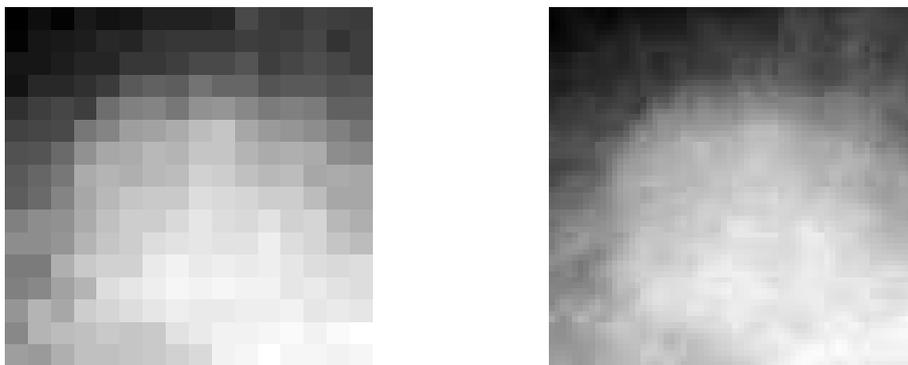
Fig. 7. Crops resizing. Resized mass (left) vs. non resized mass (right).
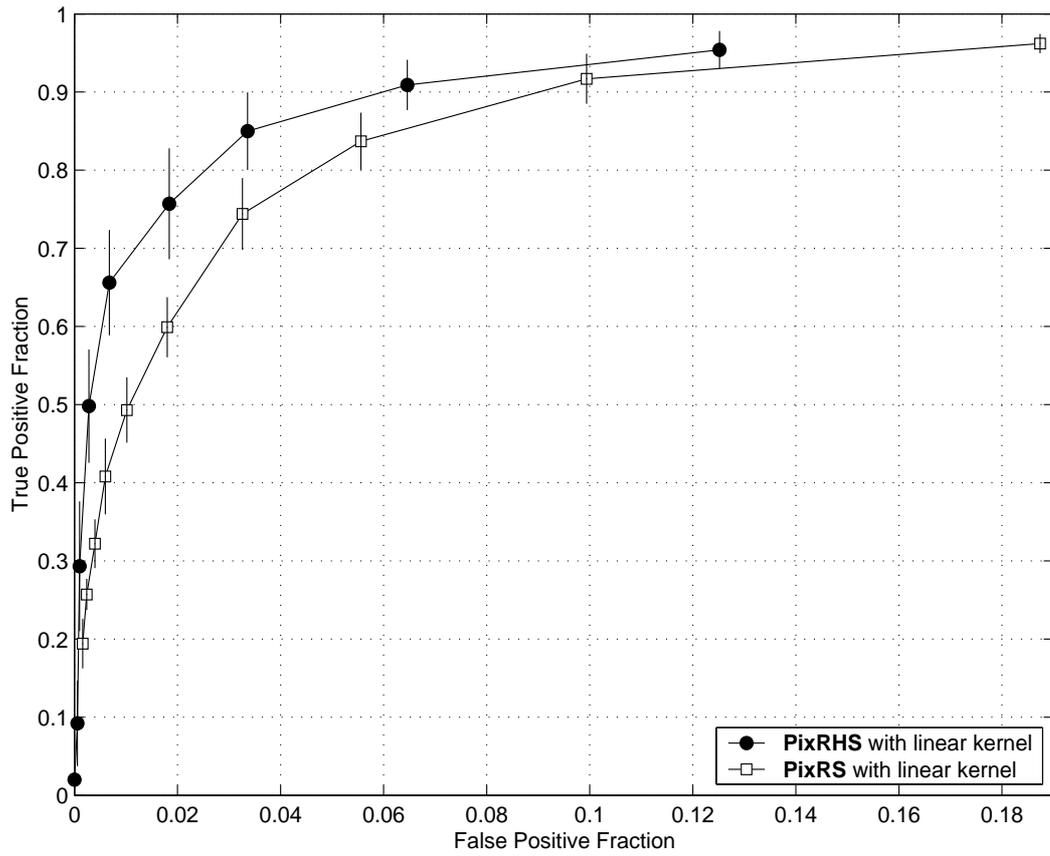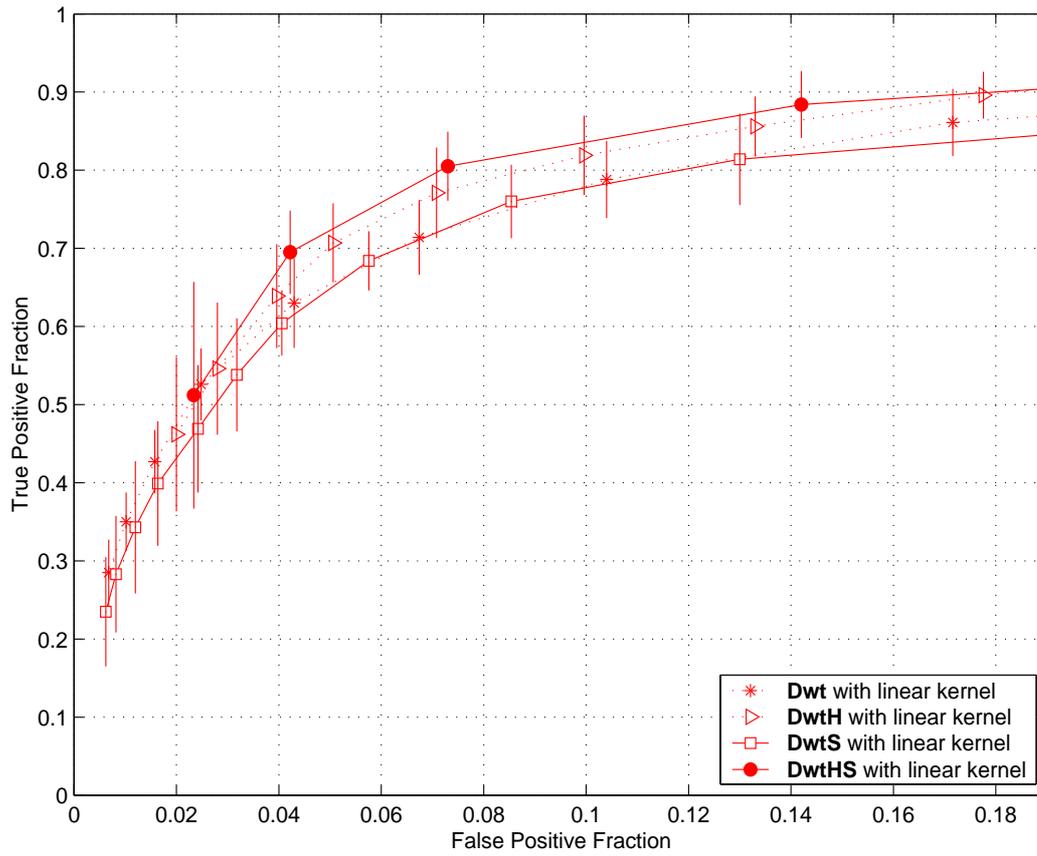
Fig. 8. Pixel–based performances using linear kernel.

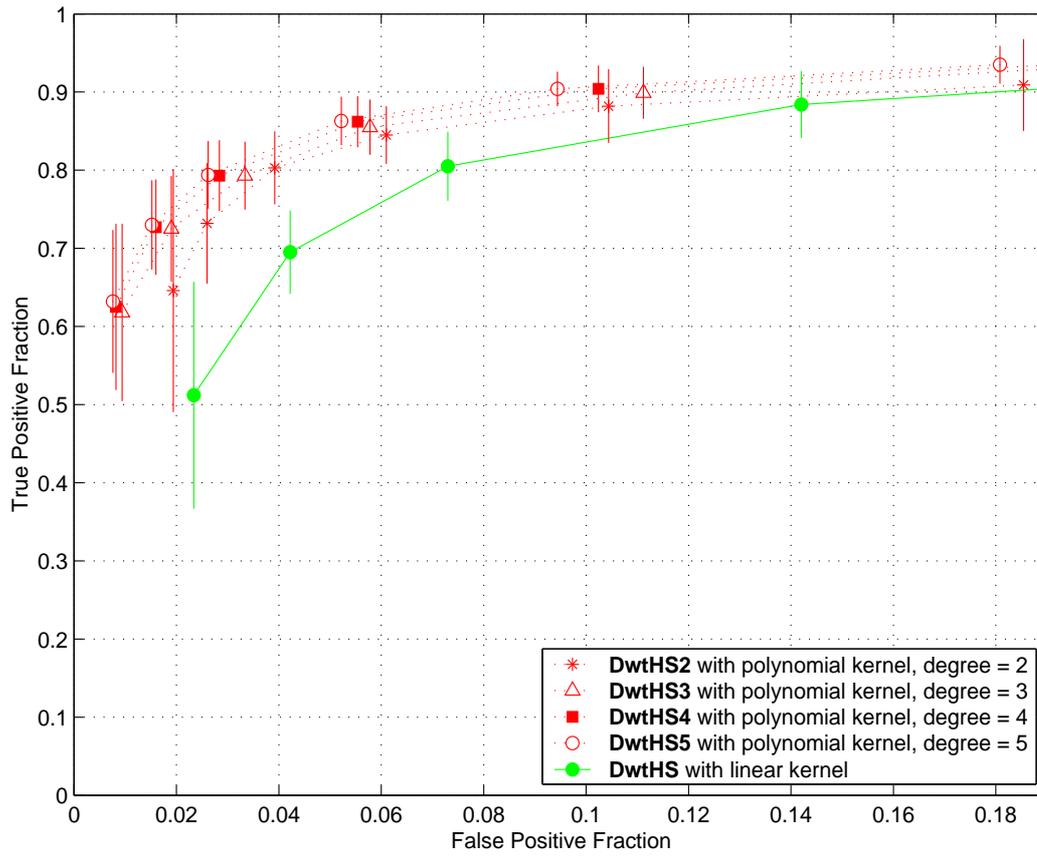Fig. 9. DWT–based performances using linear kernel.

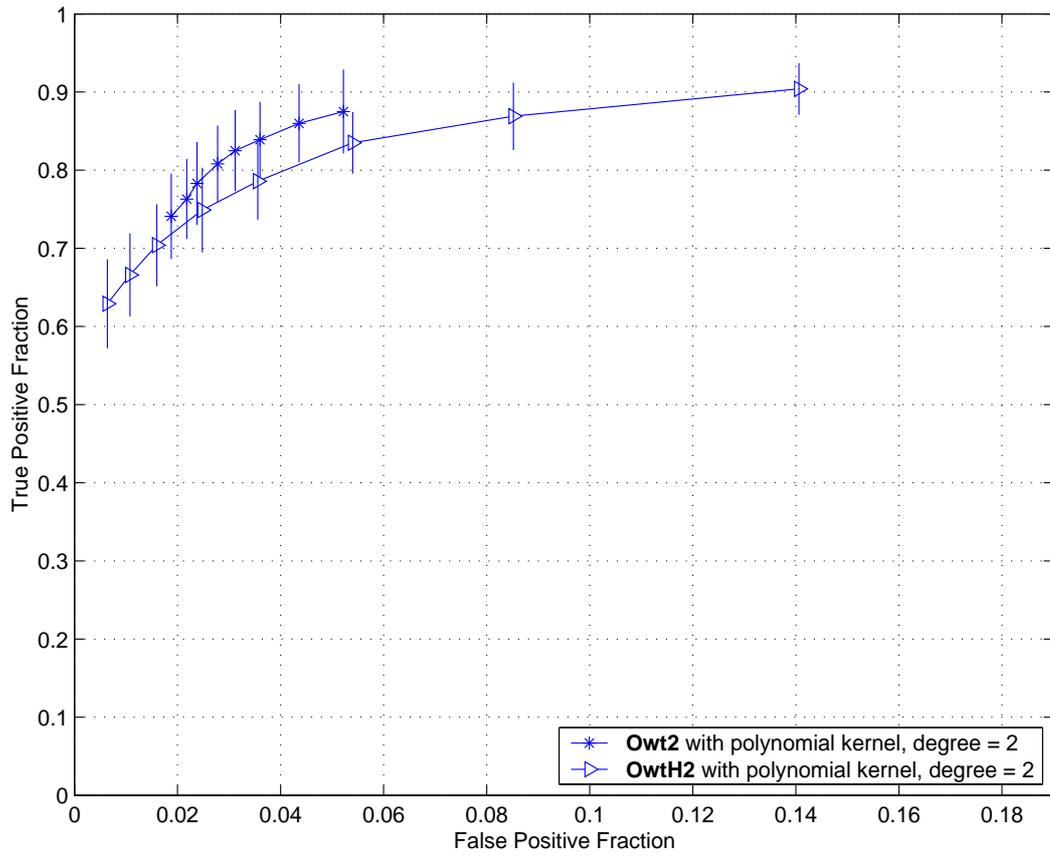Fig. 10. DWT–based performances using polynomial kernels with degree > 1.

Fig. 11. OWT–based performances using polynomial kernels with degree = 2.