A ranklet-based CAD for digital mammography

Enrico Angelini[1], Renato Campanini[1], Emiro Iampieri[1], Nico Lanconelli[1], Matteo Masotti[1], Todor Petkov[1], and Matteo Roffilli[2]

[1] Physics Department, University of Bologna, and INFN, Bologna,
Viale Berti-Pichat 6/2, 40127 Bologna, Italy
[2] Computer Science Department, University of Bologna,
Mura Anteo Zamboni 7, 40127 Bologna, Italy
nico.lanconelli@bo.infn.it

Abstract. A novel approach to the detection of masses and clustered microcalcification is presented. Lesion detection is considered as a two-class pattern recognition problem. In order to get an effective and stable representation, the detection scheme codifies the image by using a ranklet transform. The vectors of ranklet coefficients obtained are classified by means of an SVM classifier. Our approach has two main advantages. First it does not need any feature selected by the trainer. Second, it is quite stable, with respect to the image histogram. That allows us to tune the detection parameters in one database and use the trained CAD on other databases without needing any adjustment. In this paper, training is accomplished on images coming from different databases (both digitized and digital). Test results are calculated on images coming from a few FFDM Giotto Image MD clinical units. The sensitivity of our CAD system is about 85% with a false-positive rate of 0.5 marks per image.

1. Introduction

Two of the most frequent problems encountered in developing CAD systems for mammography are the following. First, the automatic detection of breast lesions can be hampered by the wide diversity of their shape, size and subtlety. Detection methods often rely on a feature extraction step: here, lesions are isolated by means of a set of characteristics. Due to the great variety of lesions, it is extremely difficult to get a common set of features effective for every kind of lesion. This is particularly true for masses, since they can vary considerably in optical density, shape, position, size and characteristics at the edge. A second difficulty arises from that the detection algorithms are often unstable, with respect to the dynamic range of the image histogram. As a matter of fact, the CAD algorithms have to be repeatedly tuned, when images coming from different systems are considered. A suitable Look Up Table (LUT) can accomplish a sort of “normalization” to the images before the CAD analysis. In this way, the same detection scheme can be applied to images coming from different detectors and acquired in different exposure conditions. Unfortunately, it is not so easy to gain a proper LUT, which can maximize the performance of the CAD for any acquisition condition.

In this paper, we present a detection system, which does not rely on any feature extraction step and which is stable with respect to the image histogram. The first attribute stems from using an SVM classifier, whilst the second derive from the ranklet representation. The algorithm automatically learns to detect the lesions by the examples presented to it. In this way, there is no *a priori* knowledge provided by the trainer: the only thing the system needs is a set of positive examples and a set of negative examples. The detection scheme codifies the image with a ranklet representation; the great amount of information handled by the algorithm is classified by means of a Support Vector Machine (SVM) classifier. SVMs have already been applied to CAD issues in mammography since 2001 [1]. An approach based on SVM classifier, without using extracted features, has been investigated both for masses and microcalcification detection [2,3,4]. Here, we present a novel use of ranklets, as an effective representation for the image crops to be classified. Ranklets are nonparametric, multiresolution and orientation selective features modeled on Haar wavelets first introduced in 2002 [5]. The first attempt to use ranklets as data representation for recognition problems was for face detection problems. Current comparative researches between wavelets and ranklets on CAD systems seem to demonstrate that ranklets are able to achieve better performances when applied to represents tumoral masses.

In this study, we validate our detection scheme with images coming from a few FFDM units: the systems used were "Giotto IMAGE MD" produced by IMS, Italy. They are based on amorphous Selenium flat panel digital detector manufactured by ANRAD Corporation, Canada. The active area of the imager is 17.4 cm × 23.9 cm with a pixel pitch of 85 micrometers; images have 2048 × 2816 pixels with 13 bit gray-level resolution. In order to have a large number of training images, we trained the CAD system both on digital images coming by the FFDM units and on digitized images coming the USF DDSM database available on the net [6].

2. Methods

The ranklet-based CAD is characterized by not requiring extracted features for detecting the breast lesions. The algorithm automatically extracts the needed information during the training phase. The CAD system has been trained to detect both clustered microcalcifications and masses. Figure 1 shows a chart of our detection scheme.

The detection scheme

The CAD detection scheme consists of two separate algorithms; one able to detect masses and another one for detecting clustered microcalcifications. The first step of the mass detection algorithm consists in a pre-selection of the suspect regions within the breast. This is achieved by means of adaptive local gray-level thresholding. All the selected pixels are then analyzed by an ensemble of three different experts. Each expert is able to accomplish a multiscale detection, in order to find out masses with size ranging from 3 mm to 35 mm. The searching performed by each expert is based

on the SVM classification of the ranklet representation of all the crops centered on the pixels selected in the first step. Finally, a region is marked as suspect mass by using a *voting* strategy on the committee of the three experts. An ensemble of experts improves the overall performance of individual experts, if the individual experts commit mistakes on different objects. Basically, a region is considered suspect only if at least two of the three experts detect that region.

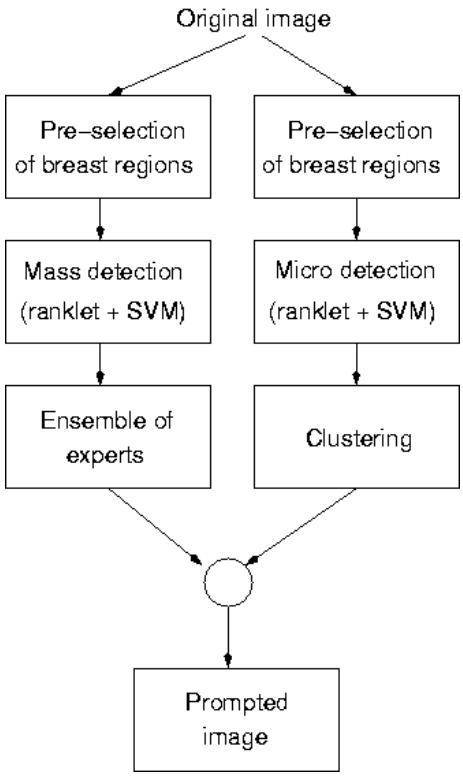

Fig. 1. Chart of the ranklet-based detection scheme.

The first step of the microcalcification detection method consists in a pre-selection of the regions containing bright spots. This is achieved by means of a statistical test calculated on a linear-filtered image. Pixels passing that test are then provided to a detector similar to the experts used for the masses. Here, a ranklet representation of the crops centered on the points extracted in the first step is obtained. After that, the crops are judged as positive or not, by using an SVM classifier. The main difference of the featureless detection between masses and microcalcifications is that in the first case a multiscale searching is used, whereas in the second case crops of fixed size are considered. The single adjacent pixels classified as suspect are the grouped together and clusterized, if more than two signals in a 1 cm^2 area are detected.

Finally, signals discovered by the masses and clustered microcalcifications detectors are joined by means of a logical OR operator, and a maximum

predetermined number of marks are presented as the final result. Signals are ranked by means of their distance from the separating hyperplane traced by SVM.

Image dataset

The training dataset consists of a number of "positive" and "negative" crops. "Positive" crops were extracted from cancer images and are centered on the lesions (masses or single microcalcifications). "Negative" crops were extracted randomly from normal images (i.e. from images without lesions). We used about 850 positive crops for training the CAD system (600 single microcalcifications and 250 opacities). A more complete description of the training procedure can be found in [2].

The dataset used for testing CAD performance consists of more than 1000 images not used for training and coming from various "Giotto Image MD" FFDM systems. Images have a pixel size equal to 85 micrometers and a gray-level resolution of 13 bits; they have been collected both in the course of the clinical evaluation of the FFDM system and subsequently during the regular clinical examinations. The database includes about 900 normal images (without lesions) and 140 images with at least one lesion, such as tumor opacities or clustered microcalcifications. The location of the lesions have been marked by expert radiologists and collected together with the images. Digital mammograms were always available in four projections per patient. Each case is relative to one patient and comprises the four projections (two cranio-caudal and two medio-lateral views). Performances are estimated by means of FROC curves, both on a per-image and a per-case basis.

False-positives marks were calculated on 154 normal images coming from screening examinations and with a follow-up of at least 1 year. These normal images were extracted from randomly chosen patients. The true positive performance were evaluated on 140 cancer images coming from symptomatic patients and confirmed by biopsy. 30 cases show masses as only signs of cancer, whereas 37 cases show only clustered microcalcifications. Three patients show both masses and microcalcifications.

The ranklet representation

Given a set of (x_1, x_2, ..., x_N) pixels, the rank transform substitutes each pixel's intensity value with its relative order (rank) among all the other pixels. This is a nonparametric transform since, given an image with N pixels, it replaces the value of each pixel with the value of its order among all the other pixels. Ranklets are designed starting from the three 2D Haar wavelets and the rank transform. In analogy to the wavelet transform, ranklet coefficients can be computed at different orientations by applying vertical, horizontal and diagonal Haar wavelet supports to each image under analysis. As a result, the orientation selectivity feature of the ranklet representation follows.

Finally, the close correspondence between the Haar wavelet transform and the ranklet transform leads directly to the extension of the latter to its multiresolution formulation. This means that, as for the wavelet transform, it is possible to compute

the ranklet transform of an image at different resolutions by means of a suitable stretch and shift of the Haar wavelet supports. At the same time, for each resolution, it is possible to characterize the image by means of orientation selective features such as the vertical, horizontal and diagonal ranklet coefficients. The multiresolution 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.

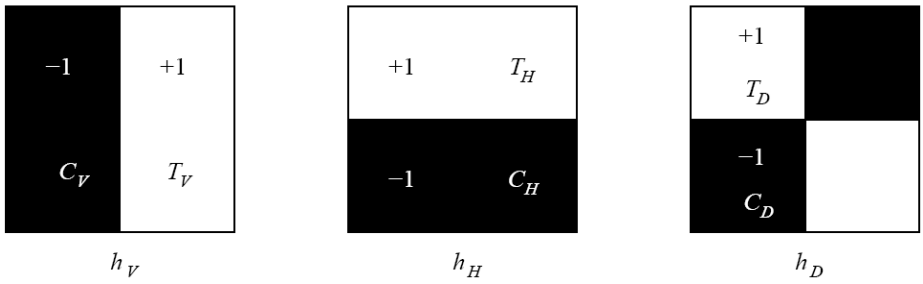

Fig. 2. The three Haar wavelet supports h_V, h_H and h_D. From left to right, the vertical, horizontal and diagonal Haar wavelet supports.

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 depicted in Fig. 2. In particular, for the vertical Haar wavelet support, the two subsets T_V and C_V are defined; similarly for the horizontal and diagonal ones. The definition of the aforementioned Haar wavelet supports forms the basis for the orientation-selective characteristic 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 $p_m \in$ T and $p_n \in$ 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.

The geometric interpretation of the so-called ranklet coefficient R_j is quite straightforward. Suppose that the image we are dealing with is characterized by a vertical edge, with the darker side on the left, where C_V is located, and the brighter side on the right, where T_V is located. R_V will be close to +1 as many pixels in T_V will have higher intensity values than the pixels in C_V. Conversely, R_V will be close to -1 if the dark and bright side are reversed. 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 use of the pixels' ranks, rather than their intensities, forms the basis for the non-parametric characteristic of the ranklet transform.

The close correspondence between the Haar wavelet transform and the ranklet transform leads directly to the extension of the latter to its multiresolution formulation. Similarly to what is done for the bidimensional Haar wavelet transform, the ranklet coefficients can be computed at different resolutions by simply stretching and shifting the Haar wavelet supports. The multiresolution 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. The possibility of computing ranklet coefficients at different resolutions forms the basis for the multiresolution characteristic of the ranklet transform.

3. Results

In order to have a remarkable number of training patterns, we accomplished the training of the CAD algorithm by using both digitized and digital images. Digitized examples were selected by cropping images from the USF DDSM database available on the net. Digital images coming from the Giotto FFDM units were used both for training and testing the CAD system. The use of images coming from various systems, without performing any normalization step has been practicable, thanks to the innate features of the ranklet transform.

The CAD system presents a sensitivity nearly equal to 85%, with a false-positive rate of 0.5 marks per image. The sensitivity has been calculated both on a per-case and on a per-image basis. In the first case, the true-positive rate is equal to the number of positive patients correctly detected over the total number of positive patients. In the latter case, results are equal to the ratio between the number of positive images correctly detected and the total number of cancer images. The false-positive rate has been computed on the normal images.

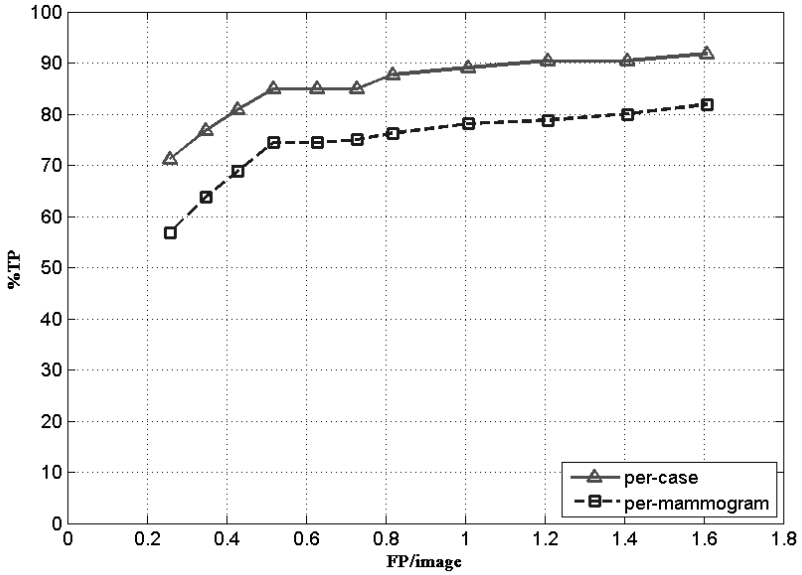

Fig. 3. FROC results of the ranklet-based CAD system on the test images. True-positive rate results are shown on a per-case and per-image basis.

Fig. 3 shows the FROC curves of our CAD system on the test images. The distinct performance for the masses and microcalcifications algorithms for a specific point of the FROC curve is the following. The masses detector shows a per-case sensitivity equal to 76% with a false-positive rate of 0.3 false-positive marks per image, whilst microcalcifications detector demonstrates a true-positive per-case rate equal to 93% with a false-positive rate of 0.2 false-positives per image.

References

1. Bazzani, A., Bevilacqua, A., Bollini, D., Brancaccio, R., Campanini, R., Lanconelli, N., Riccardi, A., Romani, D.: An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Physics in Medicine and Biology*, 46 (6), (2001) 1651-1663.
2. Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., Roffilli, M.: A novel featureless approach to mass detection in digital mammograms based on Support Vector Machines. *Physics in Medicine and Biology*, 49, (2004) 961-975.
3. El-Naqa, I., Yang, Y., Wernick, M.N., Galatsanos, N.P., Nishikawa. R.M.: A Support Vector Machine approach for detection of microcalcifications. *IEEE Trans. Med. Imag.* 21, (2002) 1552-1563.
4. Angelini, A., Campanini, R., Iampieri, E., Lanconelli, N., Masotti, M., Roffilli, M.: Testing the performances of different image representations for mass classification in digital mammograms. *International Journal of Modern Physics – C (Computational Physics and Physical Computation)*, 17 (1), (2006) 113-131.
5. Smeraldi F.: Ranklets: Orientation selective non-parametric features applied to face detection. *Proc. of the 16th International Conference on Pattern Recognition*, 3, (2002) 379–382.
6. Heat, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The digital database for screening mammography. *Digital Mammography: IWDM2000 5th Int. Workshop on Digital Mammography*, Medical Physics Publishing, (2000) 212-218.