

Exploring Cascade Classifiers for Detecting Clusters of Microcalcifications

Claudio Marrocco, Mario Molinara, and Francesco Tortorella

DAEIMI, University of Cassino
Cassino (FR), Italy

{c.marrocco,m.molinara,tortorella}@unicas.it

Abstract. The conventional approach to the detection of microcalcifications on mammographies is to employ a sliding window technique. This consists in applying a classifier function to all the subwindows contained in an image and taking each local maximum of the classifier as a possible position of a microcalcification. Although effective such an approach suffers from the high computational burden due to the huge number of subwindows contained in an image. The aim of this paper is to experimentally verify if such problem can be alleviated by a detection system which employs a cascade-based localization coupled with a clustering algorithm which exploits both the spatial coordinates of the localized regions and a confidence degree estimated on them by the final stage of the cascade. The first results obtained on a publicly available set of mammograms show that the method is promising and has large possibility of improvement.

Keywords: Microcalcifications, mammography, computer aided detection, Adaboost, cascade of classifiers, clustering.

1 Introduction

Mammography is a radiological screening technique which makes it possible to detect lesions in the breast using low doses of radiation. At present, it represents the only not invasive diagnostic technique allowing the diagnosis of a breast cancer at a very early stage, when it is still possible to successfully attack the disease with a suitable therapy. For this reason, programs of wide mass screening via mammography for the female population at risk have been carried out in many countries. A particularly meaningful visual clue of breast cancer is the presence of clusters of microcalcifications (MC), tiny granule-like deposits of calcium that appear on the mammogram as small bright spots. Their size ranges from about 0.1 mm to 0.7 mm, while their shape is sometimes irregular. Isolated MCs are not, in most cases, clinically significant. However, the low quality of mammograms and the intrinsic difficulty in detecting likely cancer signs make the analysis particularly fatiguing, especially in a mass screening where a high number of mammograms must be examined by a radiologist in a day. In this case, a Computer Aided Detection (*CADe*) could be very useful to the

radiologist both for prompting suspect cases and for helping in the diagnostic decision as a “second reading”. The goal is twofold: to improve both the sensitivity of the diagnosis, i.e. the accuracy in recognizing all the actual clusters and its specificity, i.e. the ability to avoid erroneous detections. The approach followed by the traditional CADE systems [1,2] entails a MC localization phase after that a successive clustering phase the localized regions are clustered with very simple rules based exclusively on their proximity, in order to individuate those clusters that are worth to be prompted. A widespread approach for accomplishing the localization step entails the application of statistical techniques directly to the image data [3]. To this aim, several statistical classifiers have been applied such as artificial neural network [4], support vector machine [1] or relevance vector machine [5]. All such methods rely on the sliding window technique that accomplishes the localization task by applying a classifier function on all the subwindows contained in an image and taking each local maximum of the classifier output as a pointer to a possible microcalcification. Although effective, the problem of such approaches is the computational burden involved by the huge number of subwindows typically contained in a whole image on which the classifier function (typically not very simple) should be applied.

In the Computer Vision field a very similar problem is the detection of human faces on images. In this framework, the approach presented by Viola and Jones in [6] has been particularly successful and nowadays it represents a reference solution for face detection tasks because of its fast execution and good performance. The most prominent characteristic of this algorithm is that it is still based on the sliding window technique, but the classifier function is structured as a “cascade” of simple classifiers. This allows background regions of the image to be quickly discarded while spending more computation on promising regions.

In this paper we present a method employing the cascade-based approach for detecting clustered microcalcifications. The difference with the original Viola-Jones approach is that the cascade does not make the final decision, but merely localizes on the mammogram the regions of interest (ROIs) most probable to contain a MC. The decision on the single ROI is postponed until the clustering phase, i.e. the confidence degree estimated for each ROI by the cascade is considered as an input feature for the clustering algorithm together with the spatial coordinates of the ROIs. In this way, we finally obtain a reliable partition of the ROIs in clusters of microcalcifications.

We performed several experiments on a publicly available set of mammograms and, according to the first results, the method is promising and shows large possibility of improvement.

2 The Cascade-Based Localization Method

The idea is to employ a *cascade of classifiers*, i.e., a sequence of base classifiers, to build a detector with both high specificity and sensitivity and low computational complexity. To this aim, the features employed should be both effective and simple to evaluate.

In particular, the input pattern to the detector is made of a set of rectangular features evaluated on the sliding subwindow. Figure 1 shows the features we have used for our implementation. In addition to the features employed in the original method [6] (fig. 1 a-e), we have added a further feature (fig. 1 f) more suitable for the shape of the microcalcifications. The value for each feature is calculated by subtracting the sum of the white rectangles from the sum of the black rectangles. Each of the feature types are scaled and shifted across all possible combinations on the subwindow. Since the subwindows we consider are 12x12, there are about 11000 possible features. For each of them, the evaluation can be performed very quickly thanks to a particular intermediate representation of the image, the *integral image* [6].

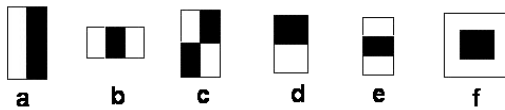


Fig. 1. The features used

Each stage of the cascade is built by means of a modified Adaboost learning algorithm that embeds a feature selection mechanism. In few words, Adaboost builds a “strong classifier” as linear combination $H(\mathbf{x}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{x})$ of “weak classifiers” $h_t(\mathbf{x})$. In the Viola-Jones approach, the weak classifier is a simple decision stump using a single feature; the peculiarity is that the features are chosen on the basis of their performance. In other words, each new weak classifier involves finding the best performing feature on all the samples of the training set: in this way, the best T features are selected and the decision about the nature of a subwindow (MC or background) can be rapidly taken.

If a given subwindow is recognized as a MC, it is passed on to the next stage, otherwise it is immediately discarded (see fig. 2). In this way, the majority of subwindows containing easily detectable background are rejected by the early stages of the detector, while the most MC-like regions go through the entire cascade. This aims at reducing the number of false positives produced by the detector.

Accordingly, the classifier $H_i(\mathbf{x})$ at a given stage is built using the samples passed through all the previous stages. While the performance desired for the whole detector is very high, the learning target for each base classifier is reasonable since each stage should guarantee a high true positive t_i rate and a sufficiently low false positive rate f_i . The performance of the whole detector with K stages will thus be given by $TPR = \prod_{i=1}^K t_i$ and $FPR = \prod_{i=1}^K f_i$. A new stage is added to the cascade until the required performance is reached. As an example, to build a detector having $TPR = 0.990$ and $FPR = 0.001$ with base

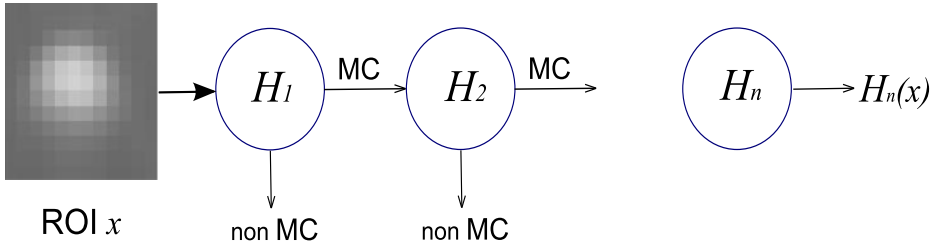


Fig. 2. The structure of the cascade

classifiers with $t_i = 0.999$ and $f_i = 0.5$, 10 stages will be needed. To this purpose, the training phase for the cascade requires a set of samples to train the single stages and a separated validation set for tuning each stage on the desired t_i and f_i . This is simply made by estimating on the validation set a threshold γ_i to be imposed on the value of $H_i(\mathbf{x})$.

It is worth noting that the cascade we employ is different from the scheme adopted in [6] which, in the final stage, provides a final decision about the ROI. In our approach, instead, the final stage produces a soft output which is a confidence degree about the presence of a MC in the ROI. In this way, the decision on the single ROI is postponed until the clustering phase so as to use for the final decision about the presence of a cluster both the estimated confidence degree and the spatial coordinates of the ROIs.

3 The Clustering Algorithm

Clustering is a well known topic in the image processing field. Roughly speaking, the goal of clustering is to achieve the best partition over a set of objects in terms of similarity. In particular, we consider a *sequential* clustering. Such kind of algorithm constructs the clusters on the basis of the order in which the points are brought and thus the final result is, usually, dependent on such order. In our system the ROIs are considered according to their confidence degrees (the largest value first) so that the clustering starts with the regions most likely to be microcalcifications. To this aim we have devised the *Moving Leader Clustering* (MLC) algorithm, which is a variation of the *leader follower clustering* [7]. It assumes as the centroid of the cluster the weighted centroid of mass of the ROIs belonging to the cluster. The centroid has to be calculated each time a new region is added to the cluster. The weight used for the ROIs is the respective confidence degree.

More formally, let us consider $n - 1$ ROIs $\mathbf{p}_1, \dots, \mathbf{p}_{n-1}$ with $n > 2$ and confidence degree respectively s_1, \dots, s_{n-1} grouped in a cluster C_1 ; its centroid is given by $\mathbf{c}_1 = \frac{\sum_1^{n-1} s_i \mathbf{p}_i}{\sum_1^{n-1} s_i}$. Let us call \mathbf{p}_n the next ROI to be considered by the algorithm and let s_n be its confidence degree. It will be added to C_1 if its distance with respect to the centroid \mathbf{c}_1 is less than a threshold R ; in this

case, the new centre of mass of C_1 will be $\mathbf{c}_1 = \frac{\sum_1^n s_i \mathbf{p}_i}{\sum_1^{n-1} s_i}$. A new cluster C_2 will be created with centroid $\mathbf{c}_1 = \mathbf{p}_n$ in the case the ROI is too far from the centroid of C_1 . An example of result of the algorithm is shown in fig. 3

In this way, the centroid of the cluster moves towards the direction where the points are more dense and with higher confidence degree and therefore, where there is a higher probability of finding new microcalcifications. When the clustering algorithm has considered all the ROIs, a post-processing operation is performed aimed at better grouping the ROIs. This consists in merging two clusters that share at least one ROI (see fig. 4). This approach has been called *Moving Leader Clustering with Merge* (MLCwM) [8] and avoids an excessive number of clusters that does not correspond to the real distributions.

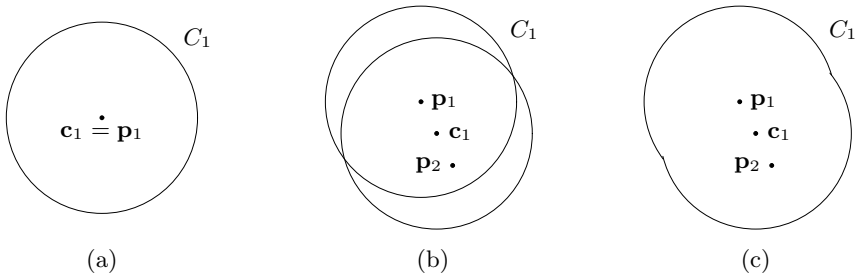


Fig. 3. The ROI \mathbf{p}_1 with highest confidence degree is the centroid \mathbf{c}_1 of a new cluster C_1 (a), when \mathbf{p}_2 is added to the cluster, the centroid moves accordingly (b) and the shape of the cluster is the union of the two circles (c)

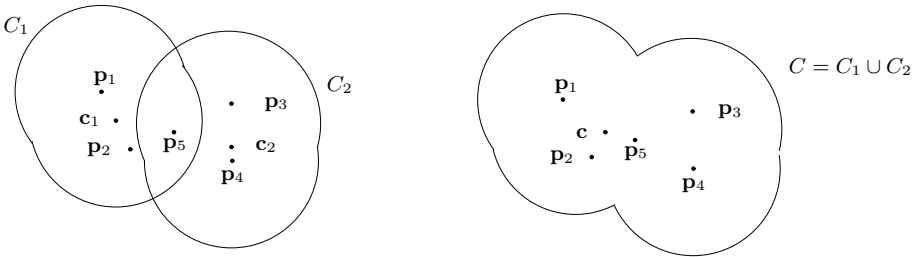


Fig. 4. When two different clusters share a ROI, they are merged and the shape of the cluster is the union of the two shapes

4 Experimental Results

The method has been tested on 145 images extracted from a publicly available database, the Digital Database for Screening Mammography (DDSM) [9]. The

employed images are relative to 58 malignant cases and contain 119 clusters and 2849 microcalcifications.

In order to build a training set for the cascade classifier, several regions corresponding to positive and negative classes have been extracted from 100 images. In particular, the positive set consisted of 186 hand labeled microcalcifications scaled and aligned to a base resolution of 12 by 12 pixels while the negative set consisted of randomly extracted images with the same base resolution. Figure 5 shows an example of such regions for both positive and negative classes. The data set so obtained has then been equally parted into a training and a validation set respectively used to train each stage of the cascade and to verify its performance. During the test phase the remaining 45 images (containing 812 microcalcifications) have been used to evaluate the performance of the proposed approach.

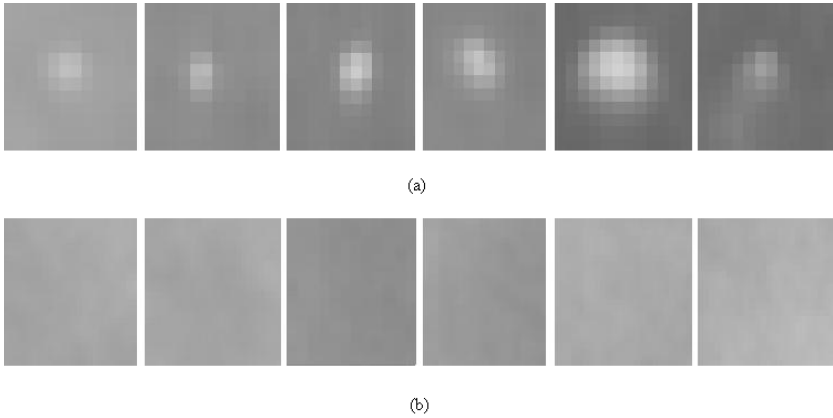


Fig. 5. An example of images used as training set: (a) positive class (microcalcifications), (b) negative class (non microcalcifications)

The cascade detector has been built by choosing a target FPR equal to 0.001 with $t_i = 0.99$ and $f_i = 0.5$. The model employed almost 6000 regions for the negative classes in training set. For the validation set, instead, the number of negative elements in the validation set was increased till almost 230000 with the aim to increase the specificity of the detector, i.e., to reduce the number of false positive at the output of the cascade classifier. The cascade we obtained consisted of 12 layers using respectively 3, 2, 2, 50, 11, 18, 7, 50, 14, 21 and 24 features. The features used in the first three layers of the cascade are shown in fig. 6. Differently from what expected, the feature shown in fig. 1.f is never used in the first three layer of the cascade, i.e., it is less discriminant than the other features in the detection of the MC regions. This is probably due to the different size of the MC that varies from image to image (as shown by the images of the positive class reported in fig. 5). The other features, instead, are able to take

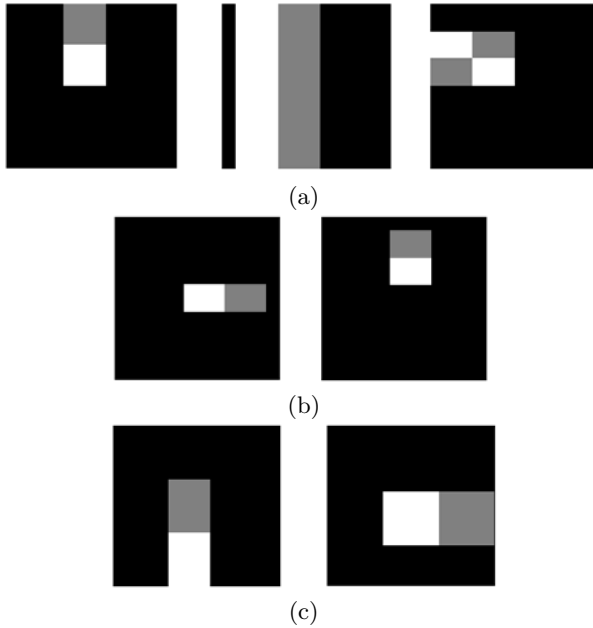


Fig. 6. The features used in the first 3 layers of the cascade

advantage of the brightness gradient in the MC training images independently of its size.

To have a comparison with other similar methods we have implemented a model of a monolithic AdaBoost classifier trained using a pixel-based approach with a 12x12 sliding window. In such case a CART decision tree with maximum depth equal to 3 and decision stumps as nodes functions has been used as weak learner. The number of boosting steps has been chosen equal to 10 while in this case the employed training set is the union of the starting training and validation set used for the cascade model.

The performance of the different proposed models have been evaluated in terms of Free-response Receiver Operating Characteristics (FROC) curve that plots the True Positive Rate (TPR), i.e., the number of regions correctly detected in the whole test set, versus the False Positive per image, i.e., the number of regions per image incorrectly classified as microcalcifications, when varying an opportune threshold. The results of the comparison are reported in fig. 7 where a threshold has been varied on the confidence degree obtained with the considered classifiers.

From this figure we can evince that the cascade approach is definitely superior to the pixel-based Adaboost in terms of both the performance measures. To remark this superiority we can note that the cascade reaches the 100% of detected clusters for 5.68 false positive per image while it is at 96% for 0.68 false positive per image. AdaBoost, instead, never reaches this performance in terms of TPR.

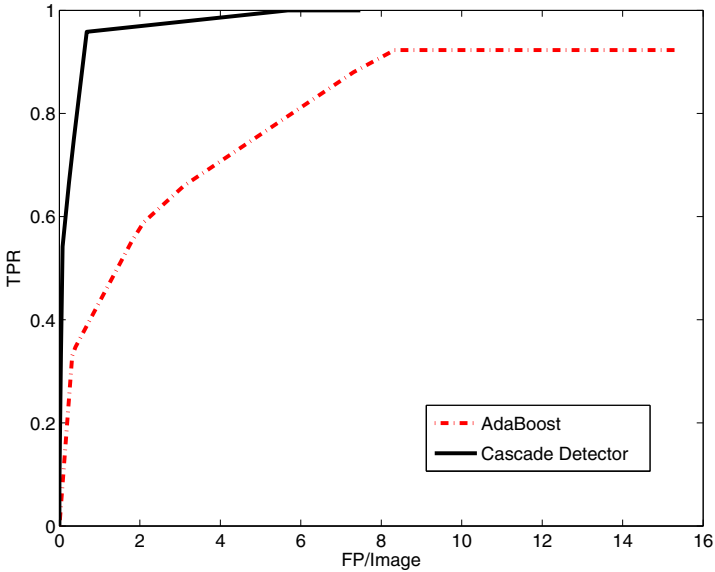


Fig. 7. The FROC curves obtained with the two compared approaches

It is also worth noting that another important advantage of the cascade approach is the considerable reduction of the elaboration time that is in terms of 64% with respect to the pixel-based AdaBoost detector.

5 Discussion and Conclusions

In this paper we have verified if the Viola-Jones approach can be profitably used in the case of detection of microcalcifications. Accordingly, we have modified this method in order to obtain as output a confidence degree that can be successively used in the clustering phase. From the first experiments, we can conclude that such approach gives good performance in both true positive rate and false positive per image. A closer look to the algorithm leads us to believe that there is still room for a significant improvement of the method performance. In fact, a critical point is the weighted accuracy used as performance measure to build the weak learners in each stage. This is not a good choice because the problem at hand is highly skewed (the number of negative samples is much greater than the positive ones) and thus a more appropriate performance index should be used (e.g. the area under the ROC curve). Another point to be considered is that, in the training phase, the boosting approach tries to improve its performance focusing on the most difficult samples and this implies a significant possibility of overfitting when dealing with very difficult samples [10]. In our case this is not very critical for the negative samples since the required FPR for each stage

is not so demanding, but it remains a problem in presence of difficult positive samples because of the high TPR to be reached. A mechanism is thus required to moderate the influence of particularly difficult positive samples which could appear during the construction of the stage. The next steps of this work will aim at modifying the original Viola-Jones approach in order to better fit the requirements for MC detection. In particular, we will focus on new features and a different tuning method for the base classifiers.

References

1. El-Naqa, I., Yang, Y., Wernick, M.N., Galatsanos, N.P., Nishikawa, R.M.: A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging* 21(12), 1552–1563 (2002)
2. Wei, L., Yang, Y., Nishikawa, R.M., Jiang, Y.: A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Transactions on Medical Imaging*, 24(3), 371–380 (2005)
3. Nishikawa, R.: Current status and future directions of computer-aided diagnosis in mammography. *Computerized Medical Imaging and Graphics* 31, 1357–1376 (2007)
4. Wu, Y., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A., Metz, C.E.: Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. *Radiology* 187, 81–87 (1993)
5. Wei, L., Yang, Y., Nishikawa, R.M., Wernick, M.N., Edwards, A.: Relevance vector machine for automatic detection of clustered microcalcifications. *IEEE Transactions on Medical Imaging* 24(10), 1278–1285 (2005)
6. Viola, P., Jones, M.: Robust real-time face detection. *International Journal of Computer Vision* 57, 137–154 (2004)
7. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. John Wiley & Sons, Chichester (2001)
8. Marrocco, C., Molinara, M., Tortorella, F.: Algorithms for detecting clusters of microcalcifications in mammograms. In: Roli, F., Vitulano, S. (eds.) *ICIAP 2005*. LNCS, vol. 3617, pp. 884–891. Springer, Heidelberg (2005)
9. Kopans, D., Moore, R., Heath, M., Bowyer, K., Philip Kegelmeyer, W.: The digital database for screening mammography. In: Yaffe, M.J. (ed.) *Proc. 5th Int. Workshop on Digital Mammography*, pp. 212–218 (2001)
10. Ratsch, G., Onoda, T., Muller, K.R.: Soft margins for adaboost. *Machine Learning* 42(3), 287–320 (2001)