

Published in final edited form as:

Phys Med Biol. 2011 January 7; 56(1): 1–17. doi:10.1088/0031-9155/56/1/001.

Detection of clustered microcalcifications using spatial point process modeling

Hao Jing¹, Yongyi Yang¹, and Robert M. Nishikawa²

Yongyi Yang: yangyo@iit.edu

¹Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 South Dearborn Street, Chicago, IL 60616, USA

²Department of Radiology, The University of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637-1463, USA

Abstract

In this work we propose a spatial point process (SPP) approach to improve the detection accuracy of clustered microcalcifications (MCs) in mammogram images. The conventional approach to MC detection has been to first detect the individual MCs in an image independently, which are subsequently grouped into clusters. Our proposed approach aims to exploit the spatial distribution among the different MCs in a mammogram image (i.e., MCs tend to appear in small clusters) directly during the detection process. We model the MCs by a marked point process (MPP) in which spatially neighboring MCs interact with each other. The MCs are then simultaneously detected through maximum *a posteriori* (MAP) estimation of the model parameters associated with the MPP process. The proposed approach was evaluated with a dataset of 141 clinical mammograms from 66 cases, and the results show that it could yield improved detection performance compared to a recently proposed SVM detector. In particular, the proposed approach achieved a sensitivity of about 90% with the FP rate at around 0.5 clusters per image, compared to about 83% for the SVM; the performance of the proposed approach was also demonstrated to be more stable over different composition of the test images.

Keywords

Clustered microcalcifications; computer-aided detection; spatial point process; marked point process

I. Introduction

Breast cancer is a common form of cancer diagnosed in women. One of the important early signs of breast cancer in mammograms is the appearance of micro-calcification (MC) clusters, which appear in 30%–50% of mammographically diagnosed cases [1]. MCs are calcium deposits of very small dimension and appear as a group of granular bright spots in a mammogram (shown in Fig. 1). Individual MCs are sometimes difficult to detect because of the surrounding breast tissue, their variation in shape and small dimension. Because of its importance in breast cancer diagnosis, accurate detection of MC clusters is an important problem. In recent years, there has been a great deal of research in development of computerized methods for automatic and accurate detection of MC clusters, which could

potentially assist radiologists in diagnosis of breast cancer. A recent review of various MC detection methods reported in the literature can be found, for example, in [2]–[4].

Most of the techniques for MC detection so far typically consist of the following two steps: first, a detection algorithm (a.k.a. pattern classifier) is employed to identify the presence of MCs in a mammogram image; next, the detected individual MCs are subsequently grouped into clusters by a clustering algorithm [5]. For example, in our previous work [6] we developed a supervised-learning approach using support vector machine (SVM) for detection of clustered MCs, which was demonstrated to outperform several well-known algorithms, e.g., the image difference technique (IDT) in [7], and the difference of Gaussian (DoG) method in [8].

In this work we develop a statistical modeling approach to further improve the accuracy of MC detection. Apart from previous reported techniques, in this approach we employ a spatial point process (SPP) to explicitly characterize the spatial clustering property of MCs in a lesion. Indeed, MCs typically appear in tightly distributed clusters when they occur in small lesions in mammograms, which are of significant interests for early cancer detection. Consequently, as explained above, a clustering step is typically employed in an MC detection algorithm in order to reduce spurious detections (false positives). This clustering property was also demonstrated to be beneficial for segmentation of MCs in a pixel-wise Markov random field (MRF) framework [9].

In our proposed approach, the individual MCs are modeled by a spatially interactive process, which are detected simultaneously via stochastic optimization. Specifically, we describe the presence of MCs in a mammogram image by a marked point process (MPP) [10], [11], which is parameterized by the number of present MCs, their spatial locations and amplitudes. We then define *a posteriori* probability model based on both image data and spatial interaction among individual MCs. The detection is achieved by maximizing this *a posteriori* probability, for which the technique of reversible jump Markov chain Monte Carlo (RJMCMC) [12], [13] is used in order to accommodate the fact that the number of MCs in a mammogram is not known *a priori*. Being a statistical sampling procedure, however, RJMCMC can be potentially expensive for numerical implementation. To address this challenge, which is great owing to the large number of pixels in a typical mammogram image, we develop strategies to greatly reduce the search for the possible constellation of the MC locations of the point process. Our results demonstrate that this approach could lead to improved detection performance when compared to the best-performing SVM detector in [6].

We note that in the literature there have been several MPP applications reported for object detection tasks, where MPP is used to characterize the underlying spatial relationship among the objects to be detected. For example, it was used for segmentation and detection of road networks [14] and buildings [15]; in medical imaging applications, it was applied for detection of brain lesions in [16] and leukocytes in [17].

The rest of the paper is organized as follows: A description of the image data model and the spatial point process model for MC cluster detection is given in Section II. Then, the RJMCMC detection algorithm is given in Section III. Issues related to our evaluation study (including mammogram data set, and performance evaluation methods) are described in Section IV. Experiment results and discussions are furnished in Section V. Conclusions and future work are given in Section VI.

II. Spatial point process model for clustered microcalcifications

In this section we begin with a description model of the image data, which consists of a superposition of MC objects in a noisy background. We then characterize the spatial distribution of MCs using a spatial point process (SPP), in which the MCs are modeled as spatially interactive objects. Afterward, we give the formulation of detection of clustered MCs as an estimation problem.

A. Data model

Given a mammogram image defined over domain Ω , we model the image function $f(\mathbf{x})$ as a superposition of a number of microcalcifications (signals) in a noisy background. That is, at pixel location $\mathbf{x} \in \Omega$, we have

$$f(\mathbf{x}) = \sum_{i=1}^N w_i K(\mathbf{x}; \mathbf{x}_i) + n(\mathbf{x}) \quad (1)$$

where $K(\mathbf{x}; \mathbf{x}_i)$ is the signal corresponding to the microcalcification located at \mathbf{x}_i , which has strength (or amplitude) w_i , $n(\mathbf{x})$ denotes the background noise, and N is the number of microcalcifications.

To model the microcalcification signal $K(\mathbf{x}; \mathbf{x}_i)$, we use a truncated Gaussian kernel centered at \mathbf{x}_i , i.e.,

$$K(\mathbf{x}; \mathbf{x}_i) = \begin{cases} e^{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}}, & \text{if } \|\mathbf{x} - \mathbf{x}_i\| \leq r_0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where r_0 is the spatial support of the signal, and σ^2 is a parameter associated with its effective size. Such a choice is based on the fact that MCs are limited in size and typically appear round in shape.

Given data $f(\mathbf{x})$, our objective is to determine the presence of the individual MCs. Our approach is to estimate the underlying parameters associated with the data model in Eq. (1), which include the number of MC objects N , their locations \mathbf{x}_i and amplitude values w_i , $i = 1, \dots, N$. For convenience, these parameters are collectively written as:

$$\Theta = \{N, \mathbf{x}_i, w_i, i = 1, \dots, N\}. \quad (3)$$

B. Spatial point process (SPP) modeling

Spatial point process is a statistical tool for modeling the spatial distribution pattern of a set of points (e.g., object locations). A marked point process (MPP) is a spatial point process in which additional labels (e.g. geometric properties of the objects) are associated with the points. That is, an MPP consists of a set of points together with their property labels $\{(\mathbf{x}_i, \mathbf{m}_i), i = 1, \dots, n\}$, where \mathbf{x}_i denote the point locations and \mathbf{m}_i denote their labels (i.e., marks). A detailed description of point process modeling can be found in [11].

We use a point process to model the spatial distribution of MC objects in a mammogram image. For convenience, let each MC object be described by $s_i = (\mathbf{x}_i, w_i)$, where \mathbf{x}_i denotes

its spatial location, and w_i denotes its amplitude. Then a set of MC objects in an image can be described by a spatial point process as $\mathbf{s} = \{s_i, i = 1, \dots, N\}$. Our goal is to exploit the spatial clustering properties among the different MC objects so as to improve their detection. For this purpose, we model this process by using a Gibbs point process, for which we define a prior distribution of the following form:

$$p(\Theta) \propto \beta^N \prod_{s_i \in \mathbf{s}} g_1(s_i) g_2(s_i, \mathbf{s}) \quad (4)$$

in which the role of each term involved is explained as follows. In Eq. (4), the occurrence of MC objects is regulated by a Poisson process, and the density parameter β is used to model the average number of MCs in an image; the term $g_1(s_i)$ is used to characterize the property of individual MC objects (specifically, their amplitude); and the term $g_2(s_i, \mathbf{s})$ is used to model the spatial interactions between an MC object and its neighboring MCs within a cluster.

Specifically, the term $g_1(s_i)$ in Eq. (4) is defined as a prior distribution on the amplitude w_i of an MC object s_i . We model $g_1(s_i)$ by a Gamma distribution, which is defined as

$$g_1(w_i; k, \theta) = \frac{1}{\theta^k \Gamma(k)} w_i^{k-1} e^{-\frac{w_i}{\theta}} \quad (5)$$

where k and θ are the distribution parameters. To demonstrate this distribution, in Fig. 2 we show a normalized histogram of the amplitude values obtained from a set of training mammograms used in our experiments (Section V.A); a Gamma distribution fitted to this histogram using ML estimation is also shown for comparison ($k = 3.59$, $\theta = 2.30$). A good agreement between the two can be observed.

Next, the interaction term $g_2(s_i, \mathbf{s})$ in Eq. (4) is defined as

$$g_2(s_i, \mathbf{s}) = g_I(s_i) \prod_{s_i \sim s_j, i < j} h(s_i, s_j) \quad (6)$$

which consists of two terms: the term $g_I(s_i)$ is used to model whether or not an MC interacts with other objects (1st order interaction), and the term $h(s_i, s_j)$ is to model its interaction with each of its neighboring MCs (2nd order interaction). Here, two objects s_i, s_j are considered neighbors (denoted by $s_i \sim s_j$) if they are within a distance d_n of each other.

The first order interaction function is defined as:

$$g_I(s_i) = \begin{cases} \alpha, & d_i < d_n \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

where d_i is the distance from object s_i to its nearest neighbor in the point process, d_n is the interaction distance. The parameter $0 < \alpha < 1$ is used to assess a penalty when an MC is isolated from any neighboring MC objects. The choice of such an interaction function is motivated by the fact that MC objects of clinical significance typically occur in small clusters. Indeed, in Fig. 3 we show a distribution of the distance between an MC and its nearest neighboring MCs in a cluster, which was obtained for all the MCs in a training set of

mammograms (Section V.A). As can be seen, an overwhelmingly majority of the MCs (97.3%) were within 60 pixels from other MCs.

The second order interaction function $h(s_i, s_j)$ is defined as

$$h(s_i, s_j) = \begin{cases} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{d_0}, & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\| < d_0 \\ 1, & \text{otherwise.} \end{cases} \quad (8)$$

In Eq. (8), d_0 is a parameter used to penalize detection of two objects that are too close to each other (e.g., even overlapping). The term $h(s_i, s_j)$ will exert a punitive interaction when the distance between two objects is less than d_0 . To illustrate this distribution, in Fig. 4 we show a histogram plot of pair-wise distances among clustered MCs obtained from a set of training mammograms used in our experiments (Section V.A). By observation, the punitive parameter in Eq. (8) was set to $d_0 = 10$, which yields a good fit between the histogram and the distribution in Eq. (8).

C. Maximum *a posteriori* (MAP) estimate

We seek a MAP solution for Θ . That is,

$$\widehat{\Theta} = \arg \max_{\Theta} \{L(\Theta) \triangleq p(\mathbf{f}|\Theta)p(\Theta)\} \quad (9)$$

where $p(\mathbf{f}|\Theta)$ denotes the likelihood function of the image data \mathbf{f} , a vector representation of the image values at all pixel locations.

We assume a Gaussian noise model with mean μ_b and variance σ_b^2 for the background noise in Eq. (1), i.e., $n(\mathbf{x}) \sim N(\mu_b, \sigma_b^2)$. Then the density function of an image sample can be written as

$$p(f(\mathbf{x})|\Theta) = (2\pi\sigma_b^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_b^2} \left[f(\mathbf{x}) - \sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) - \mu_b \right]^2 \right\}. \quad (10)$$

Note that in the above a stationary noise model is assumed for the image background. This might seem to be overly simplistic, because, as exemplified in Fig. 1, mammogram images can be rather inhomogeneous in intensity owing to many factors such as tissue structures and varying breast densities. To address this issue, in our experiments we first applied a pre-processing procedure (Section IV.B) to equalize the signal-dependent noise, the purpose being to suppress the background structures (and thereby enhancing MC signals) in a mammogram image. In Fig. 5 we show the example mammogram image in Fig. 1 after pre-processing, of which the intensity histogram is shown in Fig. 6. As can be seen, the resulting image is more uniform and amenable to the description model in Eq. (1).

Substituting both the prior model in Eq. (4) and the data model in Eq. (10) into Eq. (9), and upon algebraic manipulations, we obtain

$$\hat{\Theta} = \arg \max_{\Theta} \ln L(\Theta) = \left[- \sum_{\mathbf{x} \in \Omega} \frac{1}{2\sigma_b^2} \left[f(\mathbf{x}) - \sum_{i=1}^N w_i K(\mathbf{x}; \mathbf{x}_i) - \mu_b \right]^2 + \sum_{s_i \in \mathcal{S}} [\ln g_1(s_i) + \ln g_2(s_i, \mathbf{s})] + N \ln \beta \right]. \quad (11)$$

From Eq. (11) we observe that, for $\beta < 1$, the term $N \ln \beta$ plays the role of assessing a penalty on the number of detected objects. This can avoid the over-fitting of the data by artificially introducing too many candidate objects. In our experiments, we use β to control the level of false-positive (FP) detections. A smaller β would yield a smaller number of detections, which include both true and false detections, and vice versa. By varying β , one can obtain operating points with different FP rates.

To solve the optimization problem in Eq. (11), we use the RJMCMC method, a statistical sampling technique which allows jump between states of different dimensions. This property is useful for this problem because the number of MC objects in an image is not known beforehand. On the other hand, being a statistical sampling procedure, RJMCMC is also known to be computationally expensive. To make the matter even more challenging, a mammogram image (such as those used in our experiments) can have up to tens of millions of pixels. Thus, it becomes computationally prohibitive (and also unnecessary) if one were to sample over all these many pixel locations for the possible constellation of the MC locations in the point process. In the next section we present implementation strategies to make RJMCMC tractable for the optimization of Eq. (11). A good description of RJMCMC methods for optimization can be found in [13].

III. RJMCMC detection algorithm and implementation

In this section we first introduce the RJMCMC sampling procedure for simulating the spatial distribution of MCs in a mammogram image, and describe the sampling steps involved in the procedure. We then present our strategies to facilitate its numerical implementation. The idea is to reduce greatly the sampling space of the parameters, thereby speeding up the numerical solution.

A. RJMCMC sampling for SPP optimization

As described earlier in Section II.A, the spatial distribution of MCs in a mammogram image is modeled by a marked point process, which is characterized by the parameter set Θ in Eq. (3). In our application, the technique of RJMCMC statistical sampling is adopted to simulate this process. At each sampling step, the RJMCMC algorithm updates the parameter set Θ by randomly selecting one among a set of several so-called proposal moves. In our implementation, the following five proposal moves are used: 1) birth of a new candidate object such that it is not necessarily in interaction with any existing ones; 2) death of a candidate object which is not necessarily in interaction with any other existing ones; 3) birth of a new candidate object such that it is in interaction with some existing ones; 4) death of a candidate object which is in interaction with some existing ones; or 5) update the amplitude of an existing candidate object which is selected randomly. At each step, these five moves are selected with equal probability. The initial state for Θ is chosen to be empty, i.e., no candidate objects are included at the start.

During the sampling steps, each of these selected proposal moves is accepted with only a certain acceptance probability. Specifically, let Θ denote the current state, and Θ' denote the corresponding state after a proposal move. Let also s denote the candidate object to be added or deleted that is associated with the proposal move. Then the acceptance probability of the

proposal move is given by $r = \min\{1, R\}$, where the acceptance ratio R is given in its general form as

$$R = \frac{L(\Theta')}{L(\Theta)} \times \frac{p(\Theta' \rightarrow \Theta)}{p(\Theta \rightarrow \Theta')} \times \left| \frac{\partial \Theta'}{\partial(\Theta, s)} \right| \quad (12)$$

where $p(\Theta \rightarrow \Theta')$, $p(\Theta' \rightarrow \Theta)$ are the probabilities of the forward or its reverse moves, respectively, and $\left| \frac{\partial \Theta'}{\partial(\Theta, s)} \right|$ is the Jacobian associated with the move.

Without digression, the details of the acceptance ratio and the probability terms involved in Eq. (12) corresponding to each of the five proposal moves are given in Appendix A.

For optimization of the objective function $L(\Theta)$, the RJMCMC sampling is applied in conjunction with a simulated annealing procedure, in which the acceptance ratio in Eq. (12) is modified accordingly as:

$$R = \left(\frac{L(\Theta')}{L(\Theta)} \right)^{\frac{1}{T}} \times \frac{p(\Theta' \rightarrow \Theta)}{p(\Theta \rightarrow \Theta')} \times \left| \frac{\partial \Theta'}{\partial(\Theta, s)} \right| \quad (13)$$

where T is the temperature parameter. As $T \rightarrow 0$, the algorithm converges to the maximum of the objective function. In our experiments, we used an exponential-decaying cooling scheme, in which T was decreased exponentially with the number of iterations (Section V.C).

B. Reduction of search space for the point process

Recall that in the marked point process there are two parameters associated with each MC object $s_i = (\mathbf{x}_i, w_i)$, namely its spatial location \mathbf{x}_i and amplitude w_i . Based on discussions above, we next describe strategies for speeding up the RJMCMC procedure. First, the search for MC locations will be restricted to only a limited number of most probable candidate locations instead of over the entire image domain. Second, for an MC candidate the search for the amplitude parameter w_i will be restricted to a small neighborhood around its MAP estimate.

Search space reduction for candidate MC locations—Since MC objects are localized in size and are brighter than the background, we can first apply a simple linear detector to select only those most probable locations for further consideration. Specifically, these probable candidate locations are determined as follows:

1. Convolve the image with the MC kernel $K(\mathbf{x}; \mathbf{0})$, then detect the local maxima of the resulting output. This is essentially a template matching detector, where the kernel is used as the template.
2. Among the local maxima detected, select only those with value larger than a threshold T' . These local maxima will serve as the potential candidate locations of the MC objects. In our experiments, the threshold T' was selected such that only the top 600 candidate locations were kept for further consideration for each image. This number was chosen to be rather conservative in that it is much greater than the actual number of MCs typically seen in a mammogram image.

To further reduce the number of candidate locations, in our experiments we also first applied a pre-scouting step, of which the purpose was to exclude certain regions (such as non-tissue background) in a mammogram image from further consideration. This was based on the fact that, while a whole mammogram image can be very large in size, the spatial extent of clustered MCs is typically limited (less than 1 cm^2 in area). This pre-scouting step was applied as follows: 1) the local maxima identified in the detector output in the first step above were compared against a threshold which was set to be four times the standard deviation of the detector output of the whole image; once again, the purpose here was to identify potential MC locations; 2) the identified locations were then treated as potential MC objects, and grouped into clusters (the clustering criteria are given in Section IV.C); 3) afterward, a square region ($5\text{cm} \times 5\text{cm}$) was extracted at the center of gravity of each cluster; 4) for each image up to four such regions (with the most detected locations) were selected for further consideration. Note that, to be conservative, the region size was chosen to be much larger than that of an MC cluster above.

Search space reduction for amplitude parameters—In the general case, the amplitude parameter of a candidate MC object is initially set randomly, and subsequently updated through sampling from its distribution during the RJMCMC procedure. However, in this particular case we can obtain a closed-form estimate for it by using MAP estimation. Specifically, for a new candidate object, we set its initial amplitude parameter by directly maximizing the objective function $\ln L(\Theta)$ in Eq. (11). Consider candidate object s_{i_0} . Taking the derivative of $\ln L(\Theta)$ with respect to the amplitude parameter w_{i_0} and setting it to zero, we obtain

$$\frac{\partial \ln L(\Theta)}{\partial w_{i_0}} = \frac{1}{\sigma_b^2} \sum_{\mathbf{x} \in \Omega} [K(\mathbf{x}; x_{i_0}) (f(\mathbf{x}) - w_{i_0} K(\mathbf{x}; x_{i_0}) - \mu_b)] + \frac{(k-1)}{w_{i_0}} - \frac{1}{\theta} = 0. \quad (14)$$

Noting that $w_{i_0} > 0$, we obtain from above

$$w_{i_0} = \frac{-b + \sqrt{b^2 + 4a(k-1)\sigma_b^2}}{2a} \quad (15)$$

$$\text{where } a = \sum_{\mathbf{x} \in \Omega} K^2(\mathbf{x}; x_{i_0}), \quad b = \frac{\sigma_b^2}{\theta} - \sum_{\mathbf{x} \in \Omega} [K(\mathbf{x}; x_{i_0}) (f(\mathbf{x}) - \mu_b)].$$

Note that in Eq. (14) it is assumed that the new object s_{i_0} is not in overlap with any other candidate objects. The resulting initial estimate in Eq. (15) is dependent on the current estimate of the background statistics, i.e., μ_b, σ_b^2 , and as a consequence is likely not the final optimal solution. To accommodate this, we allow a small perturbation around this estimate for RJMCMC update, i.e., $[w_{i_0} - \Delta, w_{i_0} + \Delta]$. In our experiments, $\Delta = 1$ was used.

In our experiments, the background statistics, i.e., μ_b, σ_b^2 , were estimated from the image data. To suppress their effect, the MC objects detected at the current state were excluded from the image, and the remaining image pixels were used to estimate μ_b and σ_b^2 . Since the impact caused by the individual MCs is rather small, the estimate of μ_b and σ_b^2 was updated only for every 1,000 RJMCMC steps.

IV. Performance evaluation

A. Mammogram dataset

In this study, we used a set of 141 mammograms from 66 cases (32 cancer/34 benign) collected by the Department of Radiology at the University of Chicago. Each mammogram had one or more clusters of MCs which were histologically proven. These mammograms were digitized with a spatial resolution of 0.05 mm/pixel and 12-bit grayscale with a dimension of 3,000×5,000 pixels. The MCs in each mammogram were manually identified by a group of experienced radiologists. There were a total of 2,790 MCs in these images. These MCs were used as the ground truth in our evaluation studies.

In our experiments, the dataset was randomly partitioned into two subsets, one for training and the other for testing, each containing 33 cases. It is noted that most of the 66 cases had more than one views (mediolateral oblique view and cranio-caudal view). To avoid any potential bias, the different views of the same case were always assigned together to either the training or the testing subset, but never both.

The subset of training mammograms was used to determine all the necessary parameters associated with the detection algorithm, including the parameters associated with the model distributions introduced in Section II. The subset of testing mammogram was used exclusively for testing the detection performance; at no point was it involved in fine-tuning of the parameters of the algorithm.

B. Pre-processing of mammogram images

To suppress the effect of inhomogeneous background and intensity-dependent noise in mammograms, the images were first pre-processed with a noise equalization procedure as in [18] prior to MC detection. In addition, the non-tissue background of the mammogram excluded by using a thresholding operation on the image gradient (the background area has much lower intensity and smaller gradient amplitude than the tissue area). As an example, Fig. 3 shows the same mammogram image in Fig. 1 after pre-processing.

C. Performance evaluation using FROC and bootstrap sampling

In our experiments, the detection performance was summarized using FROC curves [19]. An FROC curve plots the correct detection rate of MC clusters (i.e., true positive fraction) versus the average number of false-positives (FP) per image over the continuum of the operating range. As explained earlier in Section II.C, in the proposed SPP method the parameter β was varied to control the FP rate; a smaller β value would lead to lower FP but also lower true detections, and vice versa.

For the FROC curves, the detected MC objects were grouped to form MC clusters using a criterion based on recommendations in [4], which was reported to yield more realistic performance than several other alternatives. Specifically, a group of objects is considered to be a cluster if the objects are connected with nearest-neighbor distances (Dnn) less than 0.5 cm and there are at least three objects within a square area of 1 cm². A detected cluster is labeled as a true positive (TP) cluster if: 1) it includes at least two true detected MCs; and 2) its center of gravity is within 1 cm of that of a known true MC cluster. Likewise, a detected cluster is labeled as a false positive (FP) cluster if: 1) it contains no true MCs, or 2) the distance between its center of gravity and that of any known cluster is larger than 1 cm.

In this study we focus primarily on the operating range when the FP rate is no more than 2 FP clusters per image, which is of main interest in clinical practice. To remove the effect of case distributions in the dataset, we conducted an FROC study using a bootstrapping

methodology [20]. A total of 2,000 bootstrap sample sets were used, based on which the area under the FROC curve (A_z) was estimated.

The bootstrapping procedure used was as follows: the subset of testing mammograms consisted of 71 images. For each bootstrap step, the same number of images was randomly sampled with replacement from the testing set; the FROC curve was obtained based on the detection performance on this bootstrap sample. By repeating this sampling procedure many times, we obtained a distribution of the bootstrap A_z values. When comparing two detection algorithms, we applied this procedure to obtain a distribution of the bootstrap difference in their performance A_z , based on which statistical inference was drawn.

V. Results and discussions

A. Determination of model parameters from training set

The proposed detection algorithm involves several parameters that need to be determined during the training phase, which include the kernel function $K(\cdot; \cdot)$ in the data model in Eq. (1) and the various parameters associated with the marked point process model. These parameters were determined as follows in our experiments.

For the MC kernel function $K(\cdot; \cdot)$, as explained earlier in Sect. II.A., a truncated Gaussian kernel as in Eq. (2) was used. To determine the kernel parameters, i.e., support radius r_0 and effect size σ , we used $K(\cdot; \cdot)$ as a template, and applied template-matching to detect the MCs in the training mammogram images. The detection results were then summarized using FROC curves, based on which the best setting for r_0 and σ was selected. In our experiments, the following different settings were considered: $r_0 = 2, 3, 4, \text{ and } 5$, and $\sigma = 1, 1.5, 2$. In the end, the following setting was selected: $r_0 = 3, \sigma = 1$. We note that the detection performance based on the training set was found rather insensitive to these parameter values. For example, $\sigma = 1$ would yield very similar performance to $\sigma = 1.5$.

For the amplitude prior $g_1(s_i)$ in Eq. (5), the amplitude value w_i was determined for each of the MCs in the training set by using maximum likelihood (ML) estimate, which amounts to least-square fitting given the Gaussian noise assumption. This resulting set of amplitude values from the training MCs was then used to determine the parameters k and θ using ML estimation for the Gamma prior in Eq. (5). The obtained distribution ($k = 3.59, \theta = 2.30$) was shown earlier in Fig. 2.

For the first-order interaction term $g_I(s_i)$ in Eq. (7), we computed the distance from an MC to its nearest neighboring MC in a cluster. Figure 3 shows the histogram of this resulting distance values for all the MCs in the training set. As can be seen, an overwhelmingly majority of the distance values (96.8%) are less than 60 pixels. Based on this observation, the interaction distance d_n in Eq. (8) was set as $d_n = 60$. The penalty parameter α in Eq. (7) was set to be the same level as β .

In our experiment, the algorithm was tested over a wide range of operating points. When the operating point was at a high FP rate, the presence of potential FP objects render the penalty in $g_I(s_i)$ in Eq. (7) ineffective once they meet the interaction distance d_n . As a safe guard against this, we imposed a requirement that the nearest neighbor should also be large enough in magnitude ($w_n > 7.3$, the median of the distribution of the magnitude values in Fig. 2).

For the second-order interaction term $h(s_i, s_j)$, we computed the pair-wise distance among MCs within clusters from the training set. A histogram plot of these distance values was shown earlier in Fig. 4. By observation, the punitive parameter in Eq. (8) was set to $d_0 = 10$.

B. Detection performance

The trained SPP model was evaluated using all the mammograms in the test subset. The detection performance is summarized using FROC results in Fig. 7, where the average FROC curve obtained from the bootstrapping procedure was shown. For comparison, the FROC result was also shown for the SVM detector previously developed in [6], which was demonstrated to have the best performance compared to several representative methods for MC detection reported in the literature.

A statistical comparison between the SPP method and SVM detector yields a mean difference of 0.105 in terms of the area under the FROC curve A_z over the FP range of [0, 2] (p -value < 0.002). This corresponds to a net increase of about 5.25% in TP fraction (sensitivity) on average over the FP range of [0, 2]; the most improvement occurs when the FP is in the range of [0, 1], averaging a net increase of about 8.71% in TP fraction. In particular, the SPP achieved a sensitivity of about 90% with the FP rate at around 0.5 clusters per image, compared to about 83% for the SVM.

Furthermore, it is also observed from the FROC results in Fig. 7 that the variation for SPP obtained from bootstrapping is smaller than that of SVM. This indicates that the performance of SPP is also more stable over the different composition of the test images.

It should be noted that the FROC curves are influenced by the criteria used in defining MC clusters. Therefore, caution should be exercised when comparing FROC results reported in the literature when they were derived using various criteria and datasets. As demonstrated in previous study [6], the FROC curves would get higher when the clustering criteria are relaxed; however, the relative ordering of the curves for the different methods were preserved. This is important for comparing different methods.

C. RJMCMC convergence

For the simulated annealing procedure the following exponential-decaying cooling scheme

was used: $T_k = T_0 \left(\frac{T_M}{T_0} \right)^{\frac{k}{M}}$, where k denotes the RJMCMC step index, M is the maximum number of steps allowed, and T_k denotes the temperature at step k . In our experiments, $M = 10^5$, $T_0 = 50$, and $T_M = 10^{-5}$ were used. The RJMCMC procedure was terminated when no state change was observed for 1,000 consecutive steps.

As an example, we show in Fig. 8 the evolution of the cooling temperature, the number of candidate objects, and the objective function over 50,000 steps of the RJMCMC procedure in a typical run. In this example, a stable state had been reached in about 30,000 steps.

The execution time of SPP depends on the searching space and the simulated annealing procedure. In our experiment, the average runtime was about 51.6 seconds per mammogram image when operated at an FP rate of 0.5 clusters per image. Our implementation was in Matlab on a 2 GHz PC with 2 GB memory.

D. Future work

In this study the size parameter σ^2 was kept uniform in the kernel function in Eq. (2) for modeling the MCs in a mammogram image. Conceivably, it might be advantageous to allow this parameter to vary so as to accommodate the size variations among different MCs, which could lead to even further improvement in detection. Furthermore, it might also be advantageous to use non-symmetrical kernel functions. This, of course, would increase the search space of the RJMCMC procedure, but is worthy of further investigation.

VI. Conclusion

In this work we developed a spatial point process (SPP) modeling approach for detection of clustered microcalcifications (MCs) in mammogram images. In this approach a marked point process (MPP) was employed to characterize the spatial distribution of clustered MCs in a mammogram image, wherein prior distributions were defined to describe both the amplitude of the MC signals and their spatial interactive patterns. The parameters associated with the prior distributions were determined from a set of training mammogram images. For a given image the detection of MC objects was through determination of the underlying point process using MAP estimation, for which a statistical sampling technique using an RJMCMC procedure was used. We presented several strategies for reducing the computational complexity of RJMCMC in implementation. For performance evaluation, we used a bootstrapping methodology to obtain the detection FROC curves. Compared to a recently reported SVM detector, the proposed SPP approach was demonstrated to achieve a net increase of about 5.25% in sensitivity on average over the FP range between 0 to 2 clusters per image (p -value <0.002). A potential drawback of the proposed approach is the extra computational load associated with the RJMCMC procedure. However, given its benefits on detection performance, it is desirable to investigate additional strategies to further reduce the computation burden of the RJMCMC sampling procedure while maintaining its detection performance.

References

1. Cancer Facts and Figures 2009. Atlanta, GA: American Cancer Society; 2009.
2. Astley SM. Computer-based detection and prompting of mammographic abnormalities. *Br J Radiol.* 2004; 77:S194–S200. [PubMed: 15677361]
3. Sampat, MP.; Markey, MK.; Bovik, AC. *Handbook of Image & Video Processing*. 2. New York: Academic; 2005. Computer-aided detection and diagnosis in mammography; p. 1195-1217.
4. Nishikawa RM. Current status and future directions of computer-aided diagnosis in mammography. *Comput Med Imaging Graphics.* 2007; 31:224–235.
5. Kallergi M, Carney GM, Gaviria J. Evaluating the performance of detection algorithms in digital mammography. *Med Phys.* 1999; 26:267–275. [PubMed: 10076985]
6. El-Naqa, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imag.* 2002; 21:1552–1563.
7. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schimidt RA. Computer aided detection of clustered microcalcifications in digital mammograms. *Med Biol Eng Comput.* 1995; 33:174–178. [PubMed: 7643656]
8. Dengler J, Behrens S, Desaga JF. Segmentation of microcalcifications in mammograms. *IEEE Trans Med Imag.* 1993; 12:634–642.
9. Karssemeijer N. Stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing.* 1992; 10(6)
10. Grayer CJ, Moeller J. Simulation procedures and likelihood inference for spatial point processes. *Scand J Statist.* 1994; 21:359–373.
11. Møller, J.; Waagepetersen, RP. *Modern statistic for spatial point processes*. presented at the 21st NordStat; Denmark. 2006.
12. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika.* Dec.1995 82:711–732.
13. Andrieu C, Freitas ND, Doucet A, Jordan MI. An introduction to MCMC for machine learning. *Machine Learning.* 2003; 50:5–43.
14. Stoica R, Descombes X, Zerubia J. A gibbs process for road extraction from remotely sensed images. *Int J Comput Vision.* 2004; 57:121–136.
15. Descombes X, Zerubia J. Marked point process in image analysis. *IEEE Signal Proc Mag.* 2002; 19:77–84.

16. Descombes X, Kruggel F, Wollny G, Gertz HJ. An object-based approach for detecting small brain lesions: application to Virchow-Robin spaces. *IEEE Trans Med Imag.* 2004; 23:246–255.
17. Gang, D.; Acton, ST. Object Identification by Marked Point Process. 39th Asilomar Conference on Signals, Systems and Computers; 2005. p. 294-297.
18. Veldkamp W, Karssemeijer N. Normalization of local contrast in mammograms. *IEEE Trans Med Imag.* 2000; 19:731–738.
19. Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free-response approach to the measurement and characterization of radiographic-observer performance. *J Appl Eng.* 4:1978.
20. Samuelson, FW.; Petrick, N. Comparing image detection algorithms using resampling. *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging;* 2006. p. 1312-1315.

Appendix A: Acceptance ratios for the different proposal moves in RJMCMC

Below we provide the probability and acceptance ratio terms associated with the five proposal moves in the RJMCMC procedure.

A. Birth or death of a free candidate object

In the proposal birth move, a candidate location \mathbf{x} is first selected with equal probability among all those candidate locations in the search space that are not already in the current state Θ ; next, an amplitude value w is drawn from a uniform distribution on the interval $[w_{\min}, w_{\max}]$ and assigned to the selected candidate location. With the optimal initial amplitudes, $w_{\min} = w_{i0} - 1$, $w_{\max} = w_{i0} + 1$ and $w_{\max} - w_{\min} = 2$. This creates a new candidate object $s = (\mathbf{x}, w)$. The new state upon the proposal move is $\Theta' = \Theta \cup \{s\}$.

The proposal probability term for this forward move is:

$$p(\Theta \rightarrow \Theta') = \frac{1}{n_1(w_{\max} - w_{\min})} \quad (\text{A.1})$$

where n_1 is the number of candidate locations in the search space that are not already in the current state Θ . On the other hand, the probability for the inverse move is:

$$p(\Theta' \rightarrow \Theta) = \frac{1}{n_2} \quad (\text{A.2})$$

where n_2 is the number of candidate objects in the state Θ' .

Moreover, the Jacobian $\left| \frac{\partial \Theta'}{\partial(\Theta, s)} \right|$ is unity for this move. Hence, the acceptance ratio for the birth move is given by:

$$R_1 = \frac{L(\Theta')}{L(\Theta)} \times \frac{n_1(w_{\max} - w_{\min})}{n_2} \quad (\text{A.3})$$

For the death move, a candidate object is randomly selected and removed with equal probability from the current state Θ . The acceptance ratio is the reciprocal of that of the birth move. That is,

$$R_2 = \frac{L(\Theta)}{L(\Theta')} \times \frac{n_2}{n_1(w_{\max} - w_{\min})} \quad (\text{A.4})$$

B. Birth or death of an interacting candidate object

Similar to above, here in the proposal birth move a candidate location \mathbf{x} is first selected with equal probability among all those candidate locations in the search space that are within the interacting distance d_n from some of the existing candidate objects in the current state Θ ; next, an initial amplitude value w is drawn from a uniform distribution on the interval $[w_{\min}, w_{\max}]$ and assigned to the selected candidate location. This creates a new candidate object $s = (\mathbf{x}, w)$. The new state upon the proposal move is $\Theta' = \Theta \cup \{s\}$.

Similarly, the acceptance ratio for the death move is given by:

$$R_3 = \frac{L(\Theta')}{L(\Theta)} \times \frac{n_3(w_{\max} - w_{\min})}{n_4} \quad (\text{A.5})$$

where n_3 is the number of candidate objects in the search space that are within the interacting distance d_n from some of the existing candidate objects in state Θ , and n_4 is the number of candidate objects which are in interaction with some of other candidate objects in state Θ' .

For the death move, the acceptance ratio is the reciprocal of that of the birth move. That is,

$$R_4 = \frac{L(\Theta)}{L(\Theta')} \times \frac{n_4}{n_3(w_{\max} - w_{\min})} \quad (\text{A.6})$$

where n_3 is the number of candidate objects in the search space that are within the interacting distance d_n from some of the existing candidate objects in state Θ' , and n_4 is the number of candidate objects which are in interaction with some of other candidate objects in state Θ .

C. Update the parameter of one existing candidate object

In this proposal move, an existing candidate object is randomly chosen among all those in the current state Θ . Then its amplitude value w is replaced with a random number drawn from a uniform distribution on the interval $[w_{\min}, w_{\max}]$. The acceptance ratio for this move is:

$$R_5 = \frac{L(\Theta')}{L(\Theta)}. \quad (\text{A.7})$$

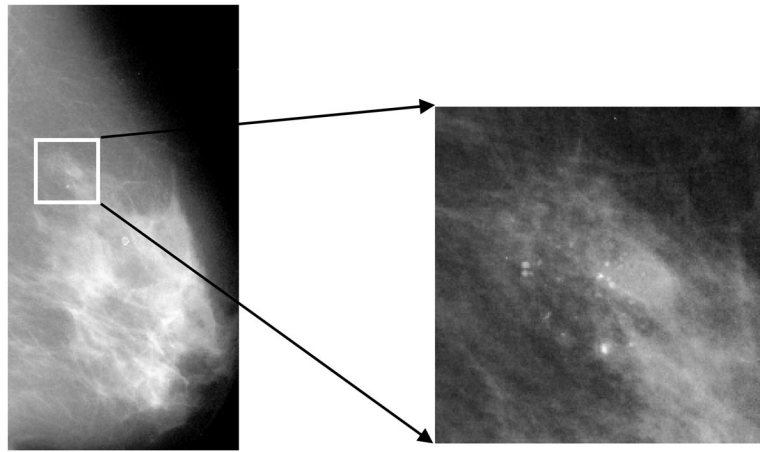


Figure 1.
A mammogram image (left) and a magnified view of a MC cluster (right).

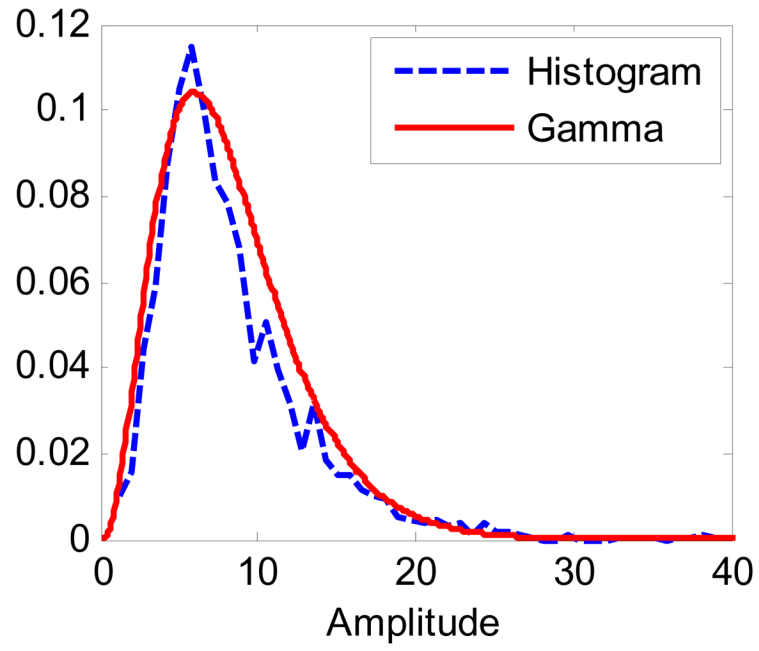


Figure 2.
Distribution of MC amplitude values.

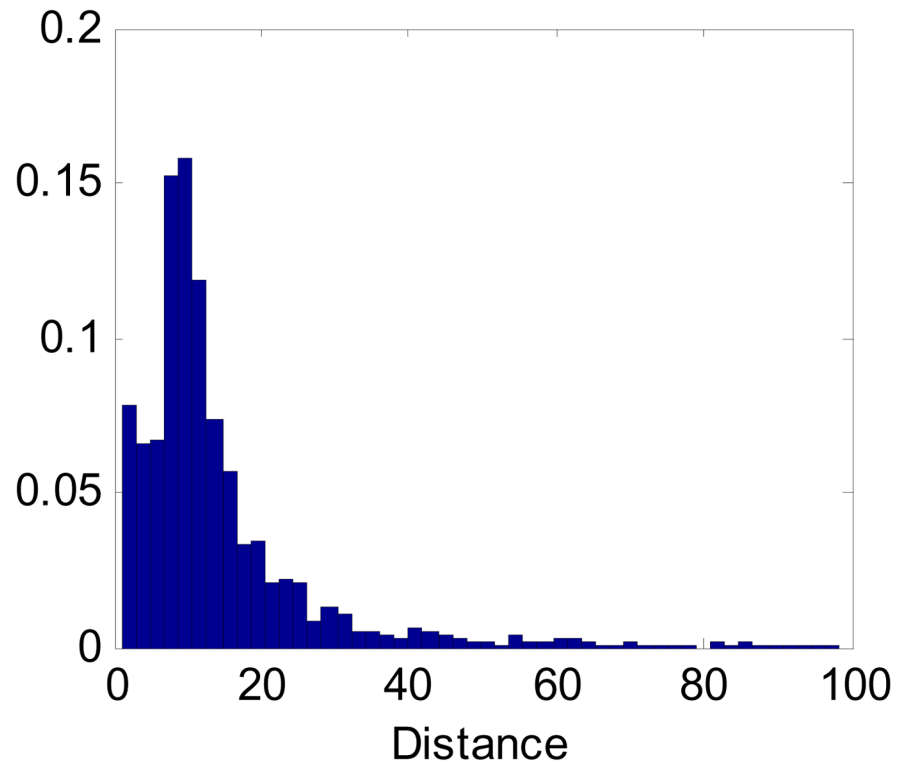


Figure 3.
Distribution of distance between an MC and its nearest neighbor.

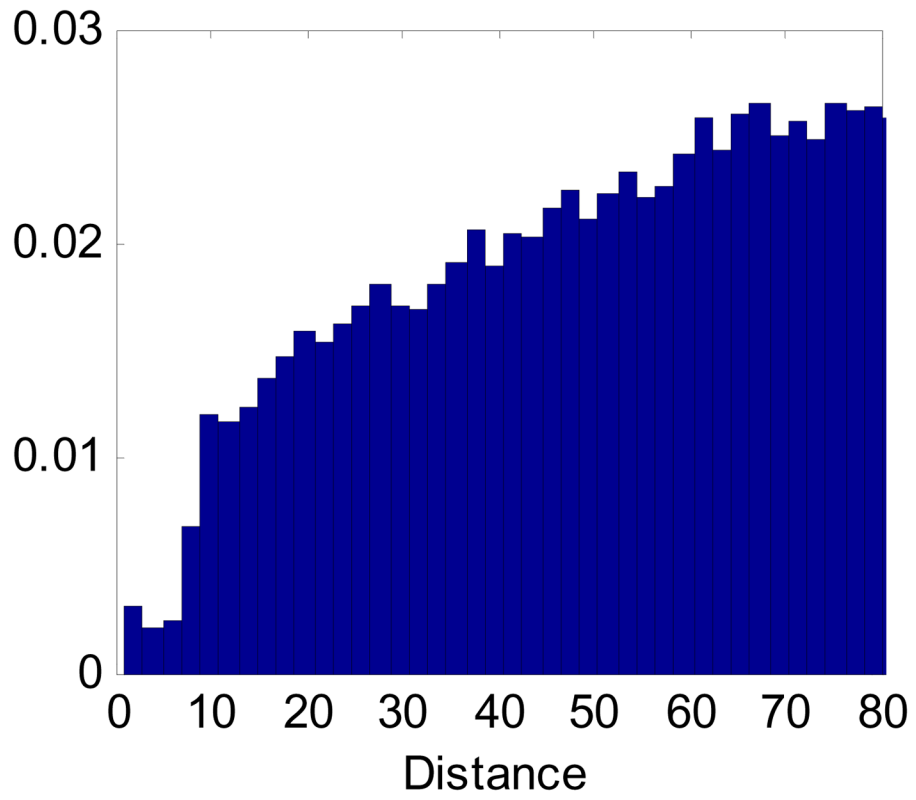


Figure 4.
Distribution of pair-wise distances between MCs in a cluster.

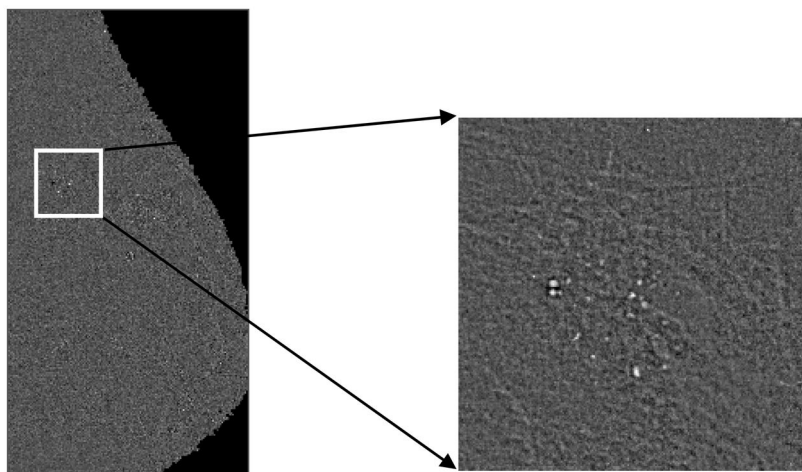


Figure 5.
Mammogram image (Fig. 1) after noise equalization.

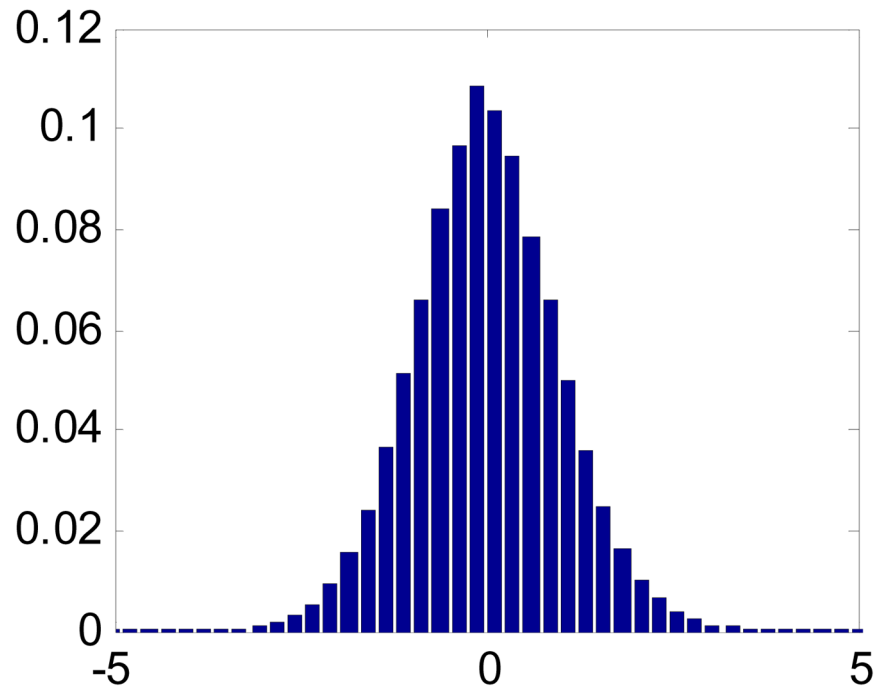


Figure 6.
Histogram of the tissue region of the image in Fig. 5.

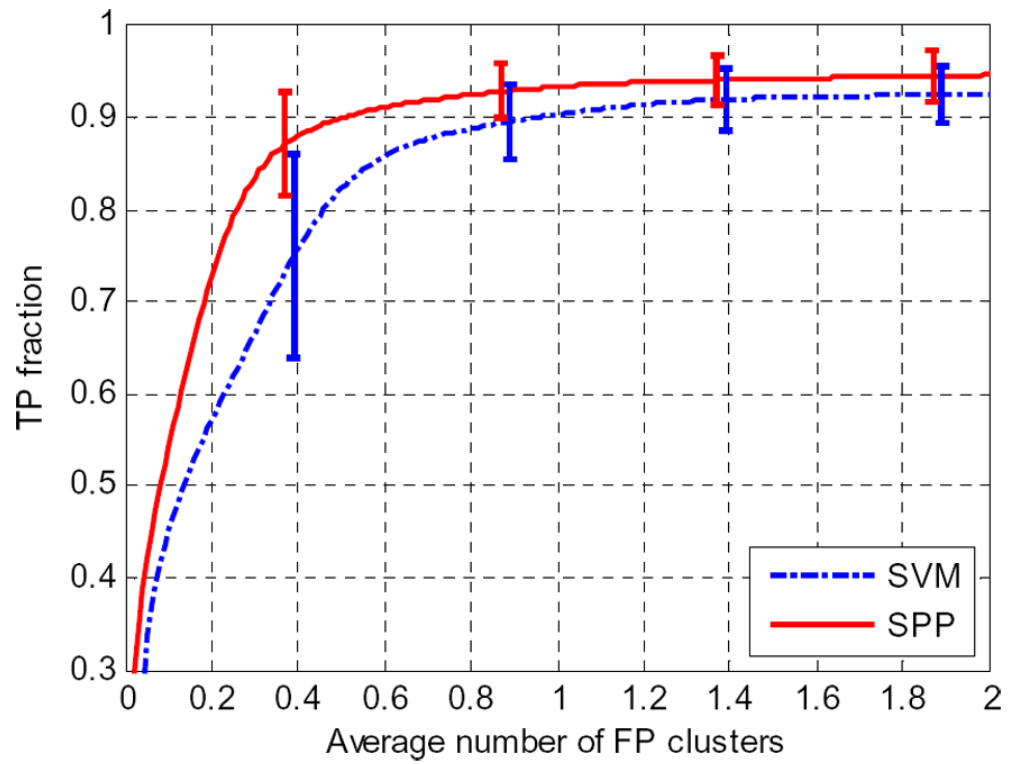


Figure 7. Average FROC curves for SVM and SPP obtained from 2,000 bootstrap samples. The most improvement by SPP occurs when the FP is in the range of [0, 1], averaging a net increase of about 8.71% in TP fraction. The smaller variation in the SPP results indicates that the performance of SPP is more stable over the different composition of the test images.

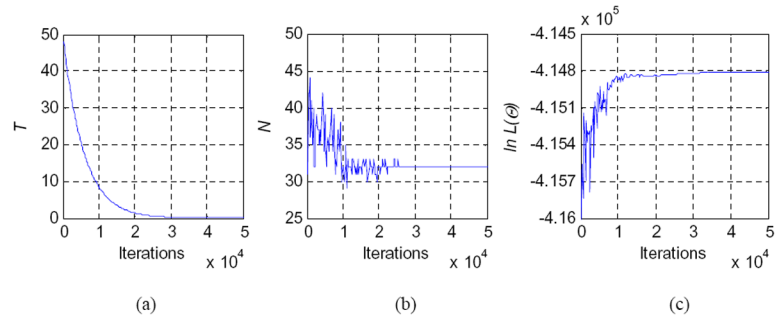


Figure 8. The evolution of (a) temperature T , (b) number of detected objects N , and (c) the value of the objective function $\ln L(\Theta)$ during the RJMCMC steps.