Stochastic and Physical Monitoring Systems
SPMS 2012
June 25 - June 30, 2012
Zlenice, Czech Republic

Evaluation of Screening Mammograms by Local Structural Mixture Models

Jiří Grim and Gobert N. Lee

Institute of Information Theory and Automation, Czech Academy of Sciences
Prague, Czech Republic

and

School of Computer Science, Engineering and Mathematics, Flinders University
Adelaide, Australia

Email: grim@utia.cas.cz & gobert.lee@flinders.edu.au

Abstract. We consider the recently proposed evaluation of screening mammograms by local statistical models. The model is defined as a joint probability density of inside grey levels of a suitably chosen search window. We approximate the model density by a mixture of Gaussian densities. Having estimated the mixture parameters we calculate at all window positions the corresponding log-likelihood values which can be displayed as grey levels at the respective window centers. The resulting log-likelihood image closely correlates with the original mammogram and emphasizes the structural details. In this paper we try to enhance the log-likelihood images by using structural mixture model capable of suppressing the influence of noisy variables.

Key words: Screening mammography; Texture information; Local statistical model; Log-likelihood image; Structural Gaussian mixture..

1 Introduction

At present the mammographic screening is a widely accepted strategy to detect early stages of breast cancer and to reduce the related mortality rates by appropriate treatment. In many countries there are numerous screening programs producing millions of mammograms to be evaluated by specially trained radiologists. Reading of mammograms is a demanding task since the early stages of pathological findings are often small, can be embedded in a complex background and show great diversity in shape, size, and location. Naturally, there is a strong motivation to help radiologists by using computer-aided decision-supporting systems but, despite of long history of research in this area, there is no satisfactory solution of the early breast cancer detection problem yet.

Recently we have proposed preprocessing of screening mammograms by using local statistical texture model [10], [11]. The idea to analyze local textural properties of mammograms comes from texture modeling [4], [5], [6]. We have found that the local properties

of homogeneous textures can be described with high accuracy by statistical dependencies between pixels of a suitably chosen small search window. In particular, at each position the window defines inside grey levels which can be written as a vector - in a fixed order. The statistical properties of grey-level variables can be expressed in full generality by the corresponding joint probability density. The underlying density can be expected to be highly multimodal and therefore we assume the model as a mixture of Gaussian densities. The mixture parameters can be estimated by EM algorithm from data obtained by scanning the mammogram with the search window.

The original motivation of the local statistical model has been to synthesize artificial textures by stepwise prediction. The quality of artificial textures is easily verified visually and therefore we have a unique possibility to "see" the quality of the statistical model. Thus, having generated an artificial texture in a reasonable quality, we can conclude that the local statistical model describes the underlying complex textural properties with sufficient accuracy. Obviously, we can use the statistical model to evaluate the local properties of the original texture image. For each position of the window we can compute the log-likelihood value which can be interpreted as a measure of typicality or novelty of the related window inside. After a series of experiments on textures (cf. [9]) we have applied this approach to screening mammograms with the aim to emphasize diagnostically important details as "unusual" locations of high "novelty" [10], [11], [12].

The log-likelihood image is a purely statistical construct without any specific relation to screening mammography. It is computed for each mammogram individually without any previous training. The log-likelihood image exactly correlates with the structural details of the source mammogram and has the same resolution. The unusual locations are emphasized by dark grey levels and suspicious areas are partly circumscribed by contour lines. We assume that, in parallel to the original screening mammogram, the log-likelihood image may be useful as an additional information to the radiologists.

In this paper we try to enhance log-likelihood images by using structural mixture model the components of which can be defined on different subspaces. By excluding the less informative noisy variables in the components we can obtain more clearly defined contour lines of suspect findings. Simultaneously, without increasing the computational complexity, we could enlarge the window to capture greater details of the mammogram. Formally, the method is applicable to monitoring systems, e.g. to identify unusual or unsafe situations in a state space.

In the following we first recall the concept of local statistical model (Sec. 2) and describe its structural modification (Sec. 3). The estimation of the structural mixture model is described in Sec. 4 and in Sec. 5 we discuss the computational experiments. Sec. 6 concludes the paper.

2 Local Statistical Model

In order to define the local statistical model we assume a suitably chosen search window. Let $\boldsymbol{x} \in \mathcal{R}^N$ be the real vector of grey levels x_n of the window inside - in a fixed arrangement

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_N) \in \mathcal{R}^N.$$

It is well known that the statistical dependencies of grey levels x_n can be described in full generality by the corresponding joint probability density $P(\boldsymbol{x})$. As the underlying probability density can be supposed to be highly multimodal we have proposed the local statistical model in the form of a Gaussian mixture (cf. [5], [6], [9], [10])

$$P(\boldsymbol{x}) = \sum_{m\in\mathcal{M}} w_m F(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m), \quad \boldsymbol{x} \in \mathcal{R}^N, \tag{1}$$

where the components are defined as products of univariate Gaussian densities:

$$F(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) = \prod_{n\in\mathcal{N}} f_n(x_n|\mu_{mn}, \sigma_{mn}), \quad \boldsymbol{x} \in \mathcal{R}^N, \tag{2}$$

$$f_n(x_n|\mu_{mn}, \sigma_{mn}) = \frac{1}{\sqrt{2\pi}\sigma_{mn}} \exp\{-\frac{(x_n-\mu_{mn})^2}{2\sigma_{mn}^2}\}. \tag{3}$$

We denote $\mathcal{M} = \{1, \ldots, M\}$, and $\mathcal{N} = \{1, \ldots, N\}$ the index sets of components and variables respectively. Recall that the total number of parameters of the mixture (1) is $(2MN + M - 1)$. The underlying diagonal covariance matrix is advantageous from the computational point of view but, moreover, it is indispensable to obtain a well interpretable "continuous" log-likelihood image (cf. Sec. 4).

In order to estimate the Gaussian mixture by EM algorithm we use the data set $\mathcal{S}$ obtained by pixelwise scanning the original mammogram with the search window

$$\mathcal{S} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots\}, \quad \boldsymbol{x}^{(i)} \in \mathcal{R}. \tag{4}$$

The corresponding log-likelihood function

$$L = \frac{1}{|\mathcal{S}|} \sum_{x\in\mathcal{S}} \log \left[\sum_{m\in\mathcal{M}} w_m F(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) \right] \tag{5}$$

can be maximized by the following EM iteration equations (cf. [6], [9])

$$q(m|\boldsymbol{x}) = \frac{w_m F(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m)}{\sum_{j\in\mathcal{M}} w_j F(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j)}, \qquad m \in \mathcal{M}, \boldsymbol{x} \in \mathcal{S}, n \in \mathcal{N}, \tag{6}$$

$$w_m^{'} = \frac{1}{|\mathcal{S}|} \sum_{x\in\mathcal{S}} q(m|\boldsymbol{x}), \qquad \mu_{mn}^{'} = \frac{1}{w_m^{'}|\mathcal{S}|} \sum_{x\in\mathcal{S}} x_n q(m|\boldsymbol{x}), \tag{7}$$

$$(\sigma_{mn}^{'})^2 = \frac{1}{w_m^{'}|\mathcal{S}|} \sum_{x\in\mathcal{S}} (x_n - \mu_{mn}^{'})^2 q(m|\boldsymbol{x}), \tag{8}$$

where the apostrophe denotes the new parameter values in each iteration.

Estimation of Gaussian mixtures in multidimensional spaces is a specific computational problem. For dimensions of order $N \approx 10^2$ the component values (2) may become too small to be correctly represented in memory and the EM algorithm may become instable. The problem can be removed if we compute the components in logarithmic form. By adding a suitably chosen constant we actually multiply the components by a coefficient which can be reduced in the fraction (6). Thus only small component values will be neglected against the current maximum (cf. [2], [9] for details).

3 Structural Mixture of Gaussian Components

In this paper we modify the original method [10] by using structural mixture model the components of which can be defined on different subspaces. The modified EM algorithm optimizes both the structural and mixture parameters simultaneously. The resulting components include only the most informative variables while the less informative "noisy" variables are replaced by a common "background". In this sense the "denoised" log-likelihood image should contain more clearly defined details.

The structural mixture model makes use of an idea originally designed for neural networks [7] and statistical recognition of binary images [8]. It can be viewed as an alternative to feature selection since, by means of a special "background" substitution technique, the computation of conditional distributions may be reduced to subsets of the most informative variables. In particular, we make substitution

$$F(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) = F(\boldsymbol{x}|\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m), \quad m \in \mathcal{M} \tag{9}$$

in (2) and introduce the structural mixture model

$$P(\boldsymbol{x}) = \sum_{m \in \mathcal{M}} w_m F(\boldsymbol{x}|\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m). \tag{10}$$

The component functions $G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m)$ include additional binary structural parameters $\phi_{mn} \in \{0, 1\}$:

$$G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m) = \prod_{n \in \mathcal{N}} \left[\frac{f_n(x_n|\mu_{mn}, \sigma_{mn})}{f_n(x_n|\mu_{0n}, \sigma_{0n})} \right]^{\phi_{mn}} \tag{11}$$

and $F(\boldsymbol{x}|0)$ is a nonzero "background" probability density usually defined as a product of unconditional univariate normal densities with the global parameters:

$$F(\boldsymbol{x}|0) = \prod_{n \in \mathcal{N}} f_n(x_n|\mu_{0n}, \sigma_{0n}), \tag{12}$$

$$\mu_{0n} = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} x_n, \quad (\sigma_{0n})^2 = -(\mu_{0n})^2 + \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} x_n^2. \tag{13}$$

Setting the structural parameter $\phi_{mn} = 0$, we can substitute any component-specific univariate distribution $f_n(x_n|\mu_{mn}, \sigma_{mn})$ by the respective univariate background distribution $f_n(x_n|\mu_{0n}, \sigma_{0n})$, i.e. we can write equivalently

$$F(\boldsymbol{x}|m) = \prod_{n \in \mathcal{N}} f_n(x_n|\mu_{mn}, \sigma_{mn})^{\phi_{mn}} f_n(x_n|\mu_{0n}, \sigma_{0n})^{1-\phi_{mn}}. \tag{14}$$

In this way the component functions $G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m)$ may be defined on different subspaces and the complexity and "structure" of the finite mixture (10) can be controlled by means of the binary parameters ϕ_{mn}. The number of involved parameters is reduced by 2 whenever $\phi_{mn} = 0$ and therefore we have the possibility to increase the dimension of the vector $\boldsymbol{x}$ without increasing the number of the estimated mixture parameters.

4 Estimation of Structural Models

Recall that the unknown mixture parameters $w_m, \mu_{mn}, \sigma_{mn}$ and also the structural parameters ϕ_{mn} can be optimized simultaneously by means of EM algorithm. Given the data set $\mathcal{S}$ we can maximize the log-likelihood function

$$L = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} \log \, [\sum_{m \in \mathcal{M}} w_m F(\boldsymbol{x}|\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m)]$$

by means of the following EM iteration Eqs.:

$$q(m|\boldsymbol{x}) = \frac{w_m G(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m, \boldsymbol{\phi}_m)}{\sum_{j \in \mathcal{M}} w_j G(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j, \boldsymbol{\phi}_j)}, \quad m \in \mathcal{M}, \boldsymbol{x} \in \mathcal{S}, n \in \mathcal{N} \tag{15}$$

$$w_m^{'} = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} q(m|\boldsymbol{x}), \quad \mu_{mn}^{'} = \frac{1}{\sum_{x \in \mathcal{S}} q(m|\boldsymbol{x})} \sum_{x \in \mathcal{S}} x_n q(m|\boldsymbol{x}) \tag{16}$$

$$(\sigma_{mn}^{'})^2 = -(\mu_{mn}^{'})^2 + \frac{1}{\sum_{x \in \mathcal{S}} q(m|\boldsymbol{x})} \sum_{x \in \mathcal{S}} x_n^2 q(m|\boldsymbol{x}), \tag{17}$$

$$\gamma_{mn}^{'} = w_m^{'} [\frac{(\mu_{mn}^{'} - \mu_{0n})^2}{\sigma_{0n}^2} + (\frac{\sigma_{mn}^{'}}{\sigma_{0n}^2})^2 - \log(\frac{\sigma_{mn}^{'}}{\sigma_{0n}})^2 - 1]. \tag{18}$$

Here the apostrophe denotes the new parameter values in each iteration.

Generally, for a fixed number of nonzero structural parameters ϕ_{mn}

$$\sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} \phi_{mn} = r, \tag{19}$$

the optimal subset of nonzero parameters $\phi_{mn}^{'}$ is defined by the r highest values $\gamma_{mn}^{'} > 0$. In this paper we have specified the fixed number of nonzero parameters ϕ_{mn} for each component separately in order to suppress the undesirably strong influence of the component weights $w_m^{'}$ in Eq. (18).

It is easily verified that the above modified iteration scheme guarantees the monotonic property of the sequence $\{L^{(t)}\}_0^{\infty}$ along with all the most important properties of EM algorithm (cf. [3]).

Let us recall that the product mixture model (1) provides log-likelihood images characterized by connected dark regions, which are partly emphasized by continuous contour lines. This property is closely related to the product components and can be explained by highly specific "topological" continuity of product mixtures. Note that a one-pixel shift of the search window generally yields a completely different data vector (despite the great overlap), because the shared gray-levels are assigned to different variables. Consequently, the likelihood values of neighboring window positions may generally differ, even by many orders. However, in case of product components, the differences of log-likelihood values are partly suppressed because the means μ_{mn} are almost uniform for any given component (cf. Fig.1). Thus the shift of the search window by one pixel actually changes only the order of product terms which are nearly the same, and therefore the product components do not differ very much in neighboring window positions.

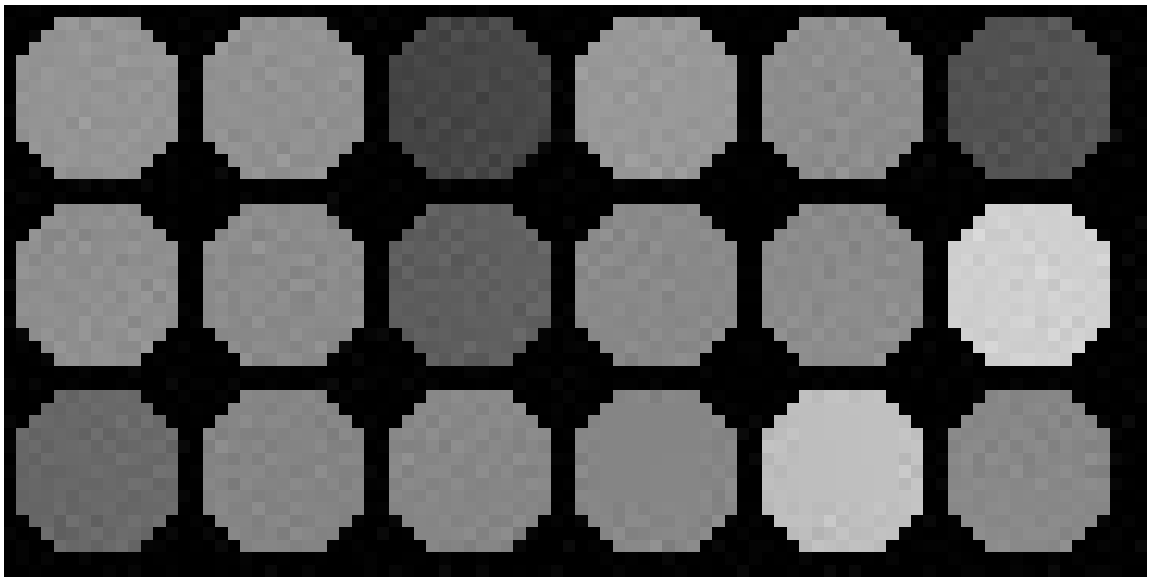

Figure 1: Examples of the component means μ_{mn} for the mammogram C-0143-1 (in the window arrangement). As it can be seen, in each component the means are nearly uniform.

5 Numerical Experiments

The local statistical model is estimated from a single mammogram, i.e. the mammograms are evaluated individually without any necessity of training by other images. For this reason the method is not confronted with the high natural variability of mammograms and with the extreme diversity of malignant findings. We apply mirror transform to right-hand-part of images to utilize the underlying symmetry and to get a more reliable background for the novelty concept.

Essentially there are two types of suspect findings - microcalcifications and masses. Isolated microcalcifications appear as small light points which can be usually identified by some kind of thresholding. The log-likelihood image identifies even one-pixel microcalcifications as dark spots having the shape of the chosen window and centered at the light pixels. The dark spots are caused by decreased likelihood values $\log P(\boldsymbol{x})$ at all positions of the window containing the "disturbing" light pixel. The shape and darkness of the spots continuously reflect the size and contrast of the underlying microcalcification.

Identification of masses is known to be more difficult than the detection of microcalcifications. Masses are of especial interest as one of the most important early signs of malignant lesions. They may be quite subtle, and may have smooth boundaries and different shapes. The recognition of masses can be facilitated by means of "contour lines" produced by local statistical model. The tendency of local statistical models to create contour lines relates to an artefact having a simple theoretical reason. The log-likelihood values $\log P(\boldsymbol{x})$ are typically "dominated" by a single component of the mixture, which is most adequate to the underlying region. A detailed numerical observation shows that the "switching" of dominating mixture components is responsible for the arising dark contour lines at the boundaries of regions having different textural properties. As the

human eye is known to be less sensitive to continuously changing contrast, the contour lines can be useful to identify the shapes of potential masses which are rather relevant from the diagnostic point of view.

In our experiments, we have used the full-field (two-view) digitized mammograms from the DDSM database of the University of South Florida [1], which have been obtained by scanning screen-film images. For the sake of model estimation, the digitized mammograms have been sub-sampled to the pixel size of about 0.1 mm. In order to facilitate the comparison of different experiments we have used a fixed square window of 13x13 pixels with trimmed corners. The corresponding dimension of data vectors was N=145 (=169-4x6), the number of components was $M = 36$ in all experiments. The mixture model has been estimated by means of EM algorithm from a data set obtained by scanning the four-view mammogram with the search window ($|\mathcal{S}| \approx 10^5 - 10^6$). The mixture parameters have been initialized randomly (cf. [10] for details), the EM iterations have been stopped by a relative increment threshold, the resulting computing time was about 2 hours.

The differences between the original and structural mixture model are illustrated by two mammograms. Fig.2 shows the medio-lateral oblique views from the screening mammograms B-3020-1 and C-0143-1 from the DDSM database. Figs. 3 and 4 illustrate the differences between the original and modified log-likelihood images for both mammograms respectively.

6 Conclusion

We propose a structural modification of a recently proposed method of diagnostic enhancement of screening mammograms. By using structural mixture model we can exclude the less informative noisy variables in the components and, as it can be seen at Figs. 3 and 4, we obtain more subtle contour lines. In comparison with the original log-likelihood image the modified image shows finer structural details and additional contour lines. Simultaneously, without increasing the computational complexity, we could increase the window size to capture larger details of the mammogram.

By considering the log-likelihood image as a final result we resign completely on any diagnostic decision-making. It is obvious that the local statistical model provides well justified and easily available information to perform e.g. segmentation of mammograms or to choose regions of interest containing suspect locations. However, we have found in many experiments that the log-likelihood image as information source is clearly preferable to any type of unique decision since it is easily interpretable, exactly correlates with the structural details of the original image and, unlike prompts or segmentation maps, does not suppress the related descriptive information.

Acknowledgement

This work has been supported by the Grant Agency of the Czech Republic - projects No. P403/12/1557 and GAČR 102/08/0593.

References

[1] K. W. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, and M. Sallam. *The digital database for screening mammography*, Kluwer Academic Publishers, 1998. (cf. *http://marathon.csee.usf.edu/Mammography/Database.html*)

[2] N. Gallego-Ortiz and Femandez-Mc-Cann, D.S. Efficient implementation of the EM algorithm for mammographic image texture analysis with multivariate Gaussian mixtures, In *Statistical Signal Processing Workshop (SSP)*, 2011 IEEE, 821–824, 2011.

[3] J. Grim. Information approach to structural optimization of probabilistic neural networks. In proceedings of: *4th System Science European Congress*, L. Ferrer et al. (Eds.), Valencia: Sociedad Espanola de Sistemas Generales, 527-540, 1999.

[4] J. Grim and M. Haindl. Texture Modelling by Discrete Distribution Mixtures, *Computational Statistics and Data Analysis*, **41**, (3–4), 603–615, 2003.

[5] M. Haindl, J. Grim, P. Somol, P. Pudil, and M. Kudo, A Gaussian mixture-based colour texture model, In *Proceedings of the 17th IAPR International Conference on Pattern Recognition*. IEEE, 177–180, Los Alamitos 2004.

[6] J. Grim, M. Haindl, P. Somol, and P. Pudil. A subspace approach to texture modelling by using Gaussian mixtures, In *Proceedings of the 18th IAPR International Conference on Pattern Recognition ICPR 2006*, **2**, 235–238, Los Alamitos, CA, USA, IEEE Computer Society, 2006.

[7] J. Grim and J. Hora. Iterative principles of recognition in probabilistic neural networks, *Neural Networks*. **21**, (6), 838-846, 2008.

[8] J. Grim, J. Kittler, P. Pudil and P. Somol. Multiple classifier fusion in probabilistic neural networks, *Pattern Analysis & Applications*, **5**, 221-233, 2002.

[9] J. Grim, P. Somol, M. Haindl, and P. Pudil. A statistical approach to local evaluation of a single texture image. In *Proceedings of the 16-th Annual Symposium PRASA 2005*, 171–176, 2005.

[10] J. Grim, P. Somol, M. Haindl, and J. Danes. Computer-aided evaluation of screening mammograms based on local texture models, *IEEE Transactions on Image Processing,*, **18**, (4), 765 –773, 2009.

[11] J. Grim. Preprocessing of Screening Mammograms Based on Local Statistical Models, In *Proc. 4th Int. Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL 2011*, Barcelona, ACM, 1–5, 2011.

[12] L. Tarassenko, P. Hayton, and N. C. et al. Novelty detection for the identification of masses in mammograms, In *Proceedings of the 4th IEE International Conference on Artificial Neural Networks*, 442447, Cambridge, UK, 1995.

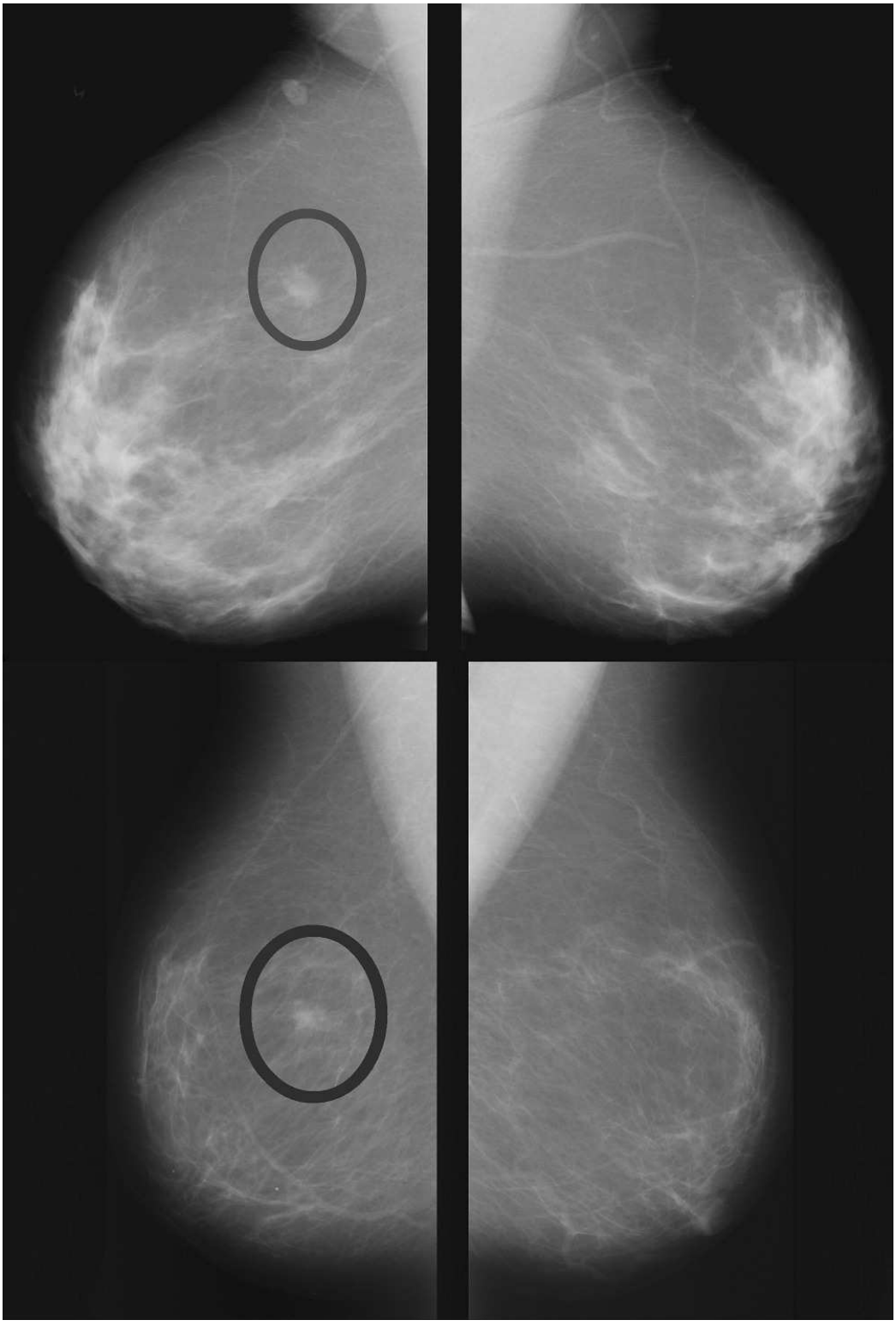

Figure 2: Original images (medio-lateral-oblique views) of the mammogram B-3020-1 (top) and C-0143-1 (bottom) from the DDSM database [1] with highlighted malignant masses.

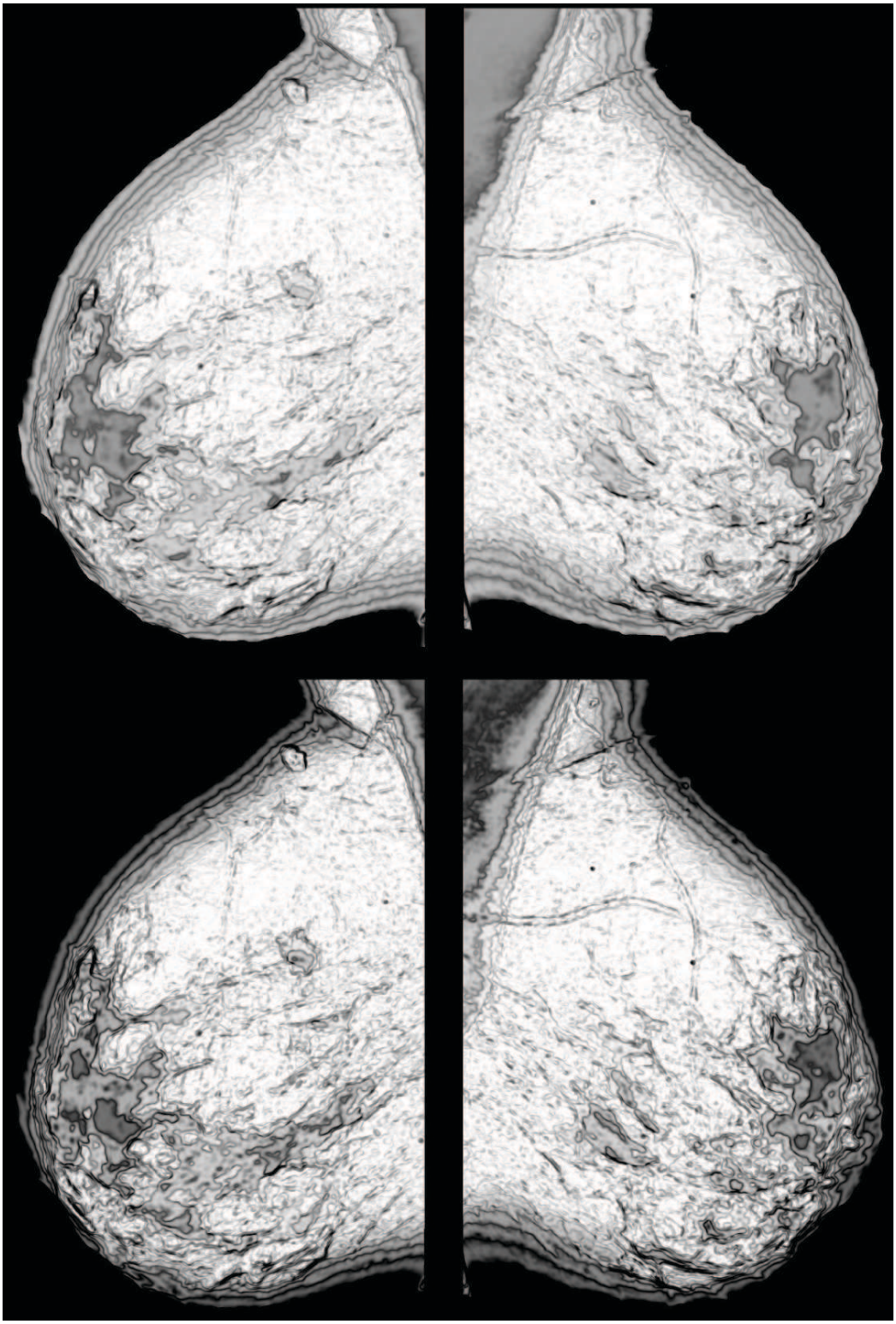

Figure 3: The log-likelihood image of the original mammogram B-3020-1 (top) in the same resolution. The grey levels are defined by the respective log-likelihood values $\log P(\boldsymbol{x})$ and the malignant mass is partly emphasized by contour lines. The "structural" log-likelihood image (bottom) contains finer details and additional contour lines.

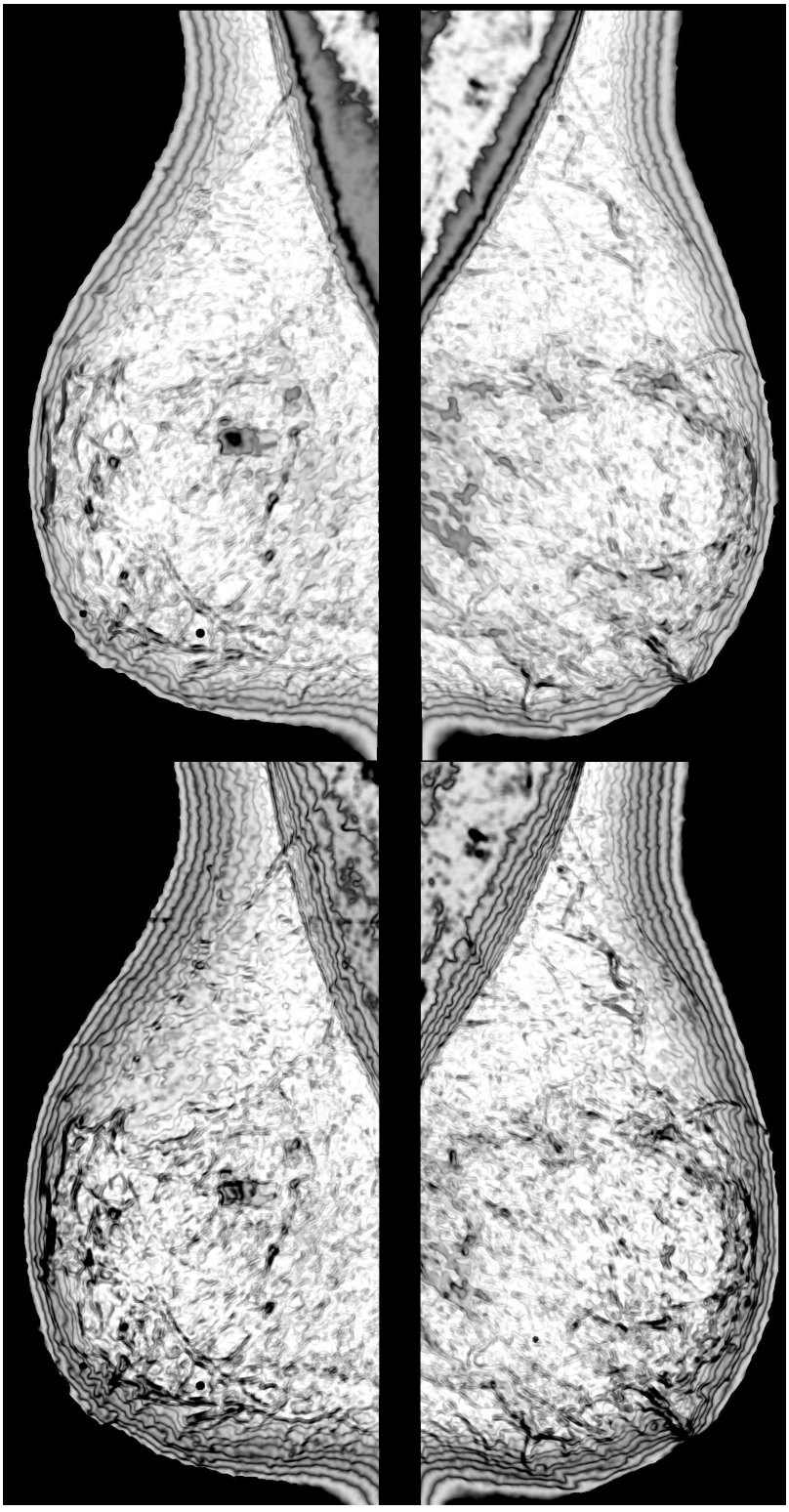

Figure 4: The log-likelihood image of the original mammogram C-0143-1 (top) in the same resolution. The "structural" log-likelihood image (bottom) contains finer details and additional contour lines.