Preprocessing of Screening Mammograms Based on Local Statistical Models

Jiří Grim

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

Department of Pattern Recognition
http://www.utia.cas.cz/RO

ISABEL 2011, October 26-29, 2011, Barcelona, Spain

Outline

Statistical Background of Mammographic Screening

Statistical Data of Breast Cancer:

- breast cancer happens to about 8% of women during their lifetime
- occurrence of malignant findings: about 1 to 3 in 1000 mammograms
- 5 to 10% of findings is proposed for surgical verification by biopsy
- about 60 to 80% of biopsies result in benign diagnoses
  ($\Rightarrow$ unnecessary physical trauma and emotional stress)
- retrospective examinations report about 10 to 20% false negative results of screening mammogram evaluation

Meaning of Mammographic Screening:

- even hardly palpable breast tumors can make metastases
- **early detection of malignant lesions** by mammographic screening is the only effective tool to decrease the breast cancer mortality rates
- $\Rightarrow$ millions of screening mammograms evaluated worldwide in one year
- $\Rightarrow$ strong motivation for computer-aided evaluation

Local Evaluation of Screening Mammograms

Idea of the log-likelihood image: LITERATURE

to emphasize mammographic lesions as "untypical" locations of high "novelty" and facilitate diagnostic evaluation of screening mammograms

local properties: square search window with cut-off corners
$\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathcal{X}, \quad x_n \;\approx\;$ grey-levels of the window inside

local statistical model: multivariate probability density $P(\mathbf{x})$
$\log P(\mathbf{x}) \;\approx\;$ measure of typicality of the window inside $\mathbf{x}$

METHOD: approximation of the density $P(\mathbf{x})$ by Gaussian mixture

- **data set:** by scanning the mammogram with the search window
- **EM algorithm:** to estimate the Gaussian mixture $P(\mathbf{x})$
- **log-likelihood image:** $\log P(\mathbf{x})$ displayed as grey-levels at window center
- **interpretation:** dark grey-levels indicate "suspect" locations

Estimation of Local Statistical Model

Gaussian mixture of product components:

$$P(\mathbf{x}) = \sum_{m=1}^{M} w_m F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) = \sum_{m=1}^{M} w_m \prod_{n=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_{mn}} \exp\{-\frac{(x_n - \mu_{mn})^2}{2\sigma_{mn}^2}\} \right]$$

log-likelihood function: (data set $\mathcal{S}$ by scanning the image)

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

EM algorithm: (initial parameters chosen randomly)

$$q(m|\mathbf{x}) = \frac{w_m F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m)}{\sum_{j \in \mathcal{M}} w_j F(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j)}, \quad w_m^{'} = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} q(m|\mathbf{x})$$

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

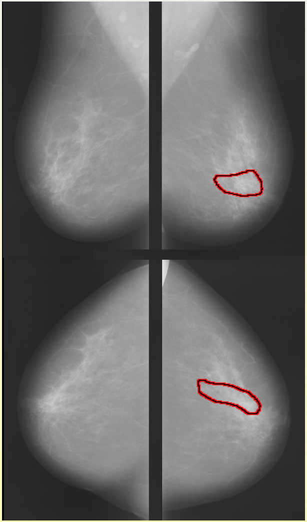
Computational Experiments

- **local statistical model:** estimated from a single mammogram
- $\Rightarrow$ each mammogram is evaluated individually
- $\Rightarrow$ the method need not be trained by other images
- $\Rightarrow$ it is not confronted with high natural variability of mammograms
- mirror transform is applied to right-hand-part of images to utilize the underlying symmetry

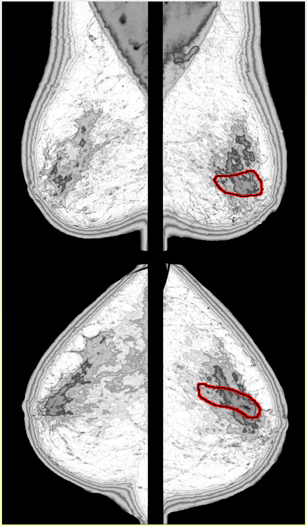
- **source images:** four-view digitalized mammograms: two medio-lateral and two cranio-caudal images
- **source database:** South Florida University http://marathon.csee.usf.edu/Mammography/Database.html
- **search window:** square window of 13 x 13 pixels with cut-off corners, dimension of **x** is $N = 145$ $(= 169 - 4 \times 6)$
- **data set:** by scanning the four-view mammogram with the search window $(|\mathcal{S}| \approx 10^5 - 10^6)$

C-0002-1: pleomorphic calcif., segmentally distributed

original image

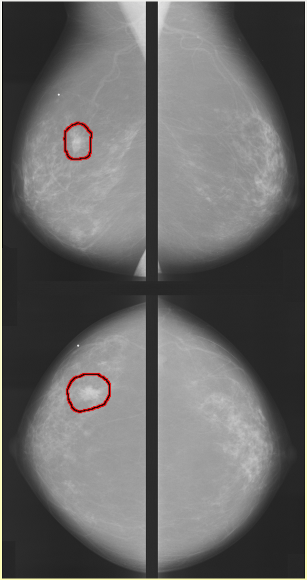


log-likelihood image

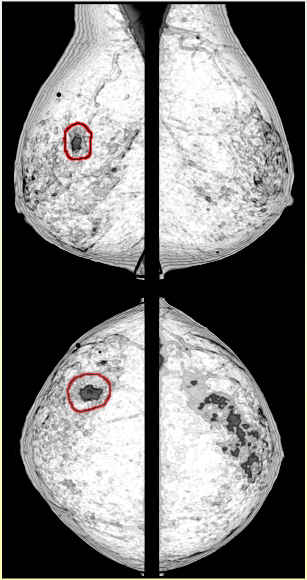


C-0016-1: segmentally distributed calcification

original image

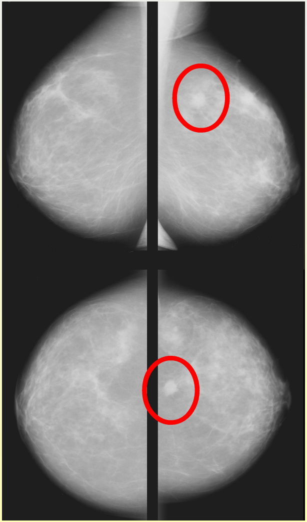


log-likelihood image

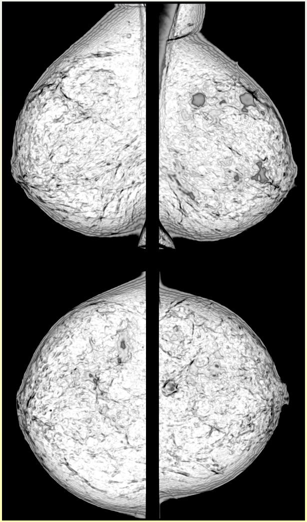


B-3017-1: Mass, lobulated-architectural distortion

original image

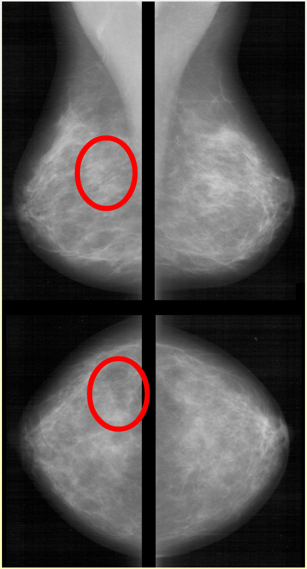


log-likelihood image

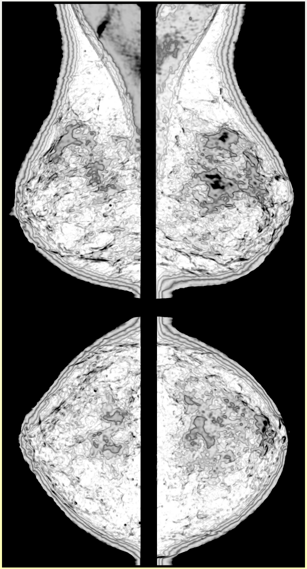


D-4163-1: punctate calcification, clustered distribution

original image



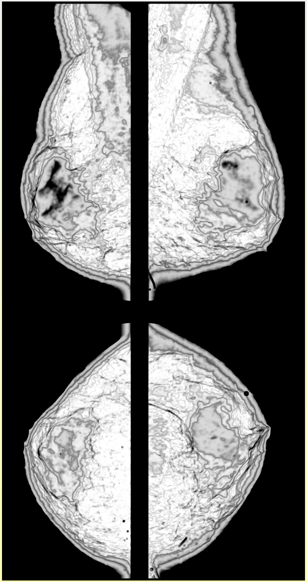
log-likelihood image



B-3056-1: mass, focal-asymmetric density, margins n/a

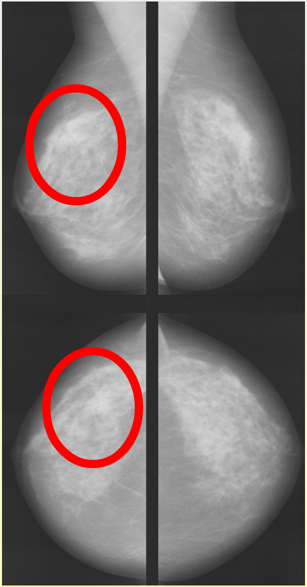
original image

log-likelihood image

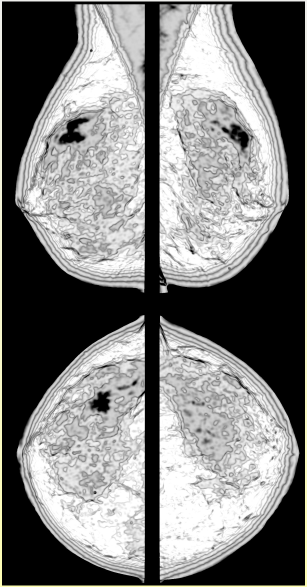


C-0001-1: mass, irregular shape, spiculated margins

original image

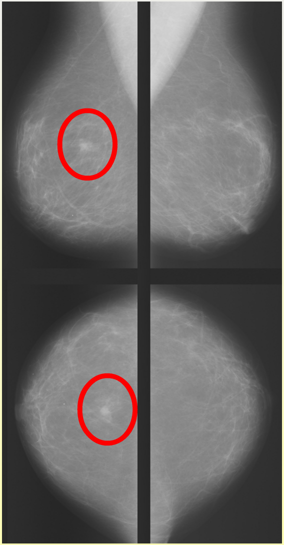


log-likelihood image

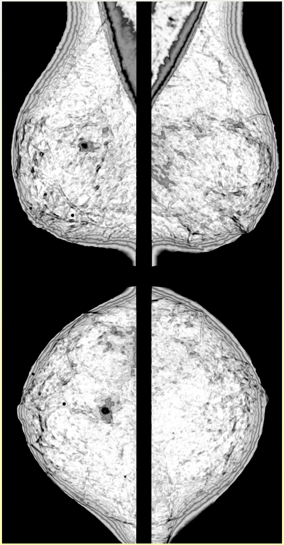


C-0143-1: mass, irregular shape, ill-defined margins

original image

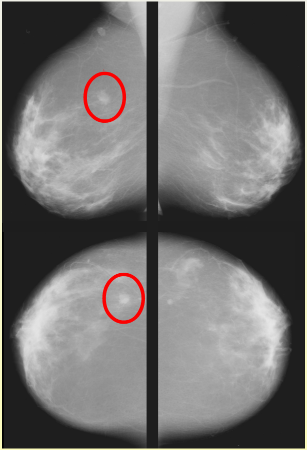


log-likelihood image

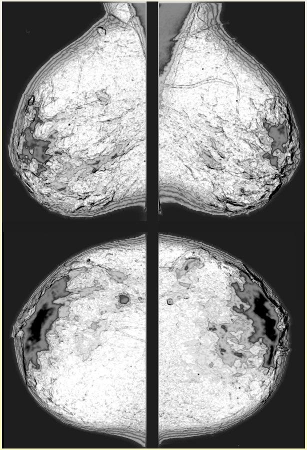


B-3020-1: mass, lobulated shape, ill defined margins

original image



log-likelihood image



Weighted Modification of Local Statistical Model

GOAL: to suppress benign findings and increase sensitivity

$\gamma(\mathbf{x}) \approx$ **weighting function**

$$\gamma(\mathbf{x}) \approx \frac{1}{N}\sum_{n=1}^{N} x_n, \quad \sum_{x\in\mathcal{S}} \gamma(\mathbf{x}) = 1,$$

weighted log-likelihood function:

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

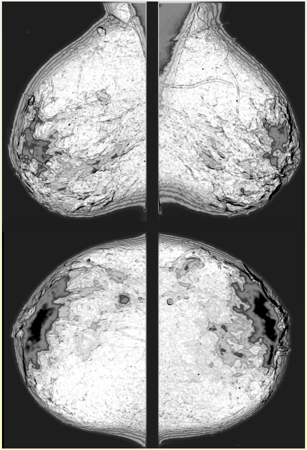
modified EM algorithm $\Rightarrow$ weighted log-likelihood image

$$q(m|\mathbf{x}) = \frac{w_m F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m)}{\sum_{j\in\mathcal{M}} w_j F(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j)}, \quad w_m^{'} = \sum_{x\in\mathcal{S}} \gamma(\mathbf{x}) q(m|\mathbf{x})$$

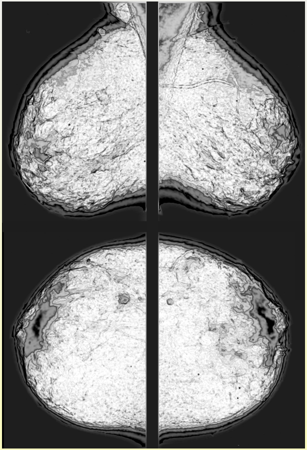
$$\mu_{mn}^{'} = \frac{1}{w_m^{'}} \sum_{x\in\mathcal{S}} x_n \gamma(\mathbf{x}) q(m|\mathbf{x}), \qquad (\sigma_{mn}^{'})^2 = \frac{1}{w_m^{'}} \sum_{x\in\mathcal{S}} (x_n - \mu_{mn}^{'})^2 \gamma(\mathbf{x}) q(m|\mathbf{x})$$

B-3020-1: mass, lobulated shape, ill defined margins

log-likelihood image



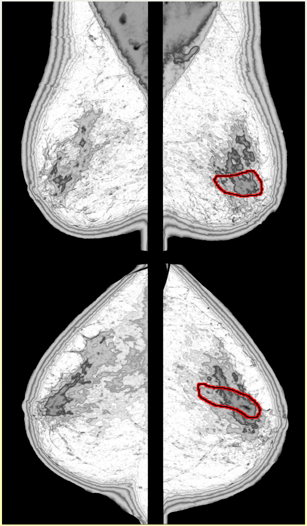
weighted log-likelihood image



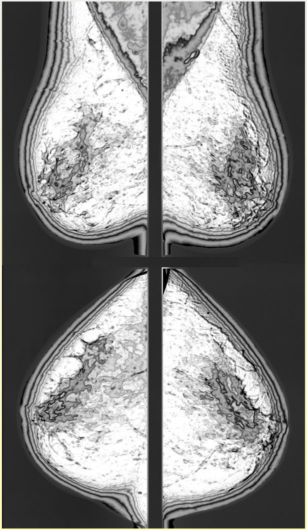
Remark: Weighted modification (right) contains finer details and provides additional contour-lines in the light regions

C-0002-1: pleomorphic calcif., segmentally distributed

log-likelihood image

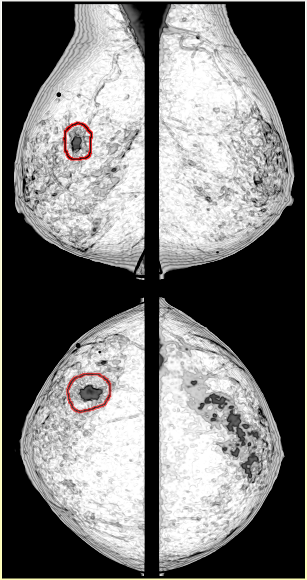


weighted log-likelihood image

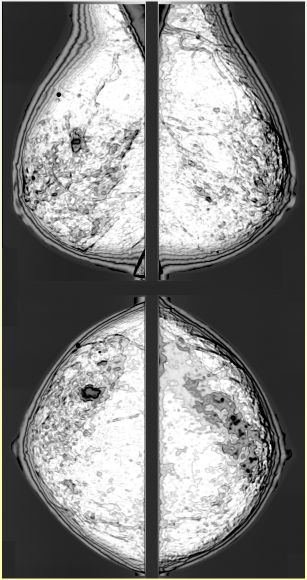


C-0016-1: segmentally distributed calcification

log-likelihood image

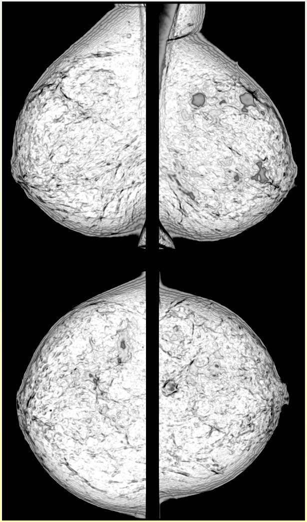


weighted log-likelihood image

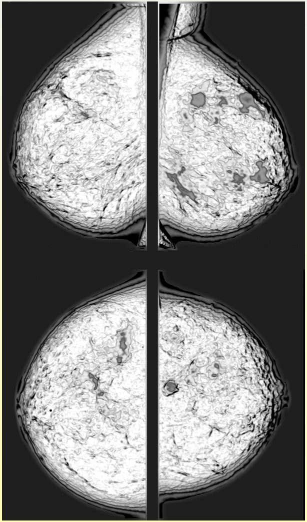


B-3017-1: Mass, lobulated-architectural distortion

log-likelihood image

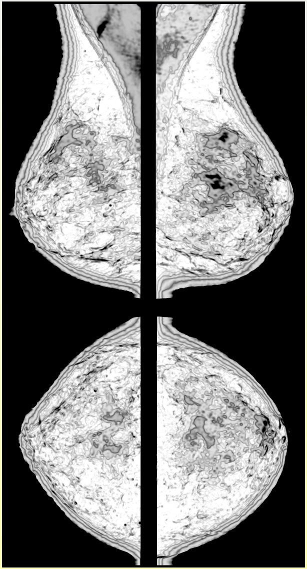


weighted log-likelihood image

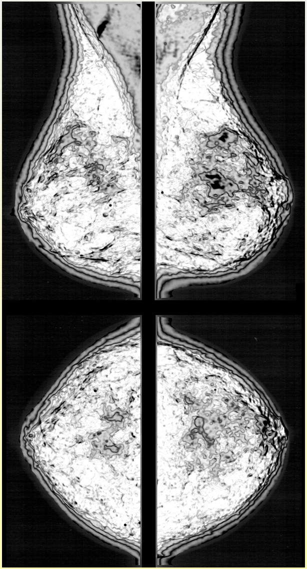


D-4163-1: punctate calcification, clustered distribution

log-likelihood image

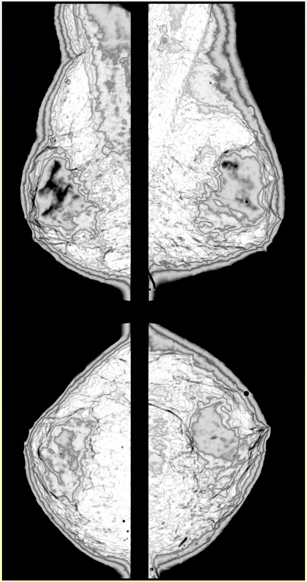


weighted log-likelihood image

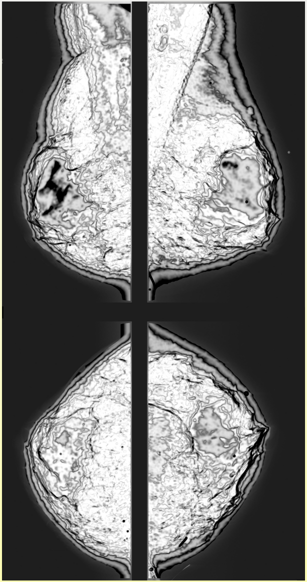


B-3056-1: mass, focal-asymmetric density, margins n/a

log-likelihood image

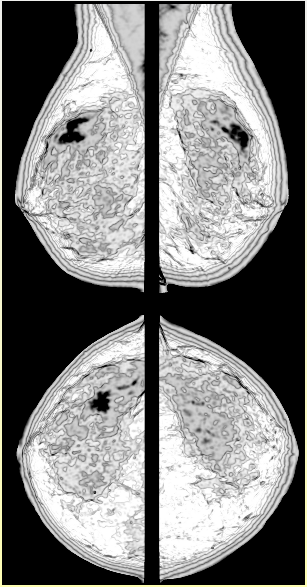


weighted log-likelihood image

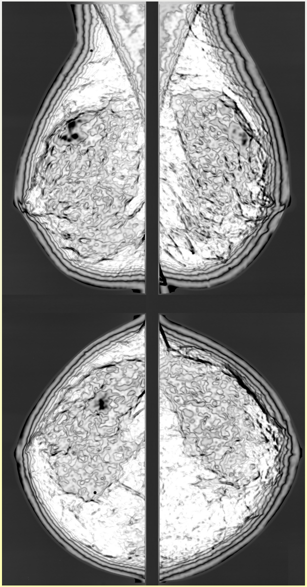


C-0001-1: : mass, irregular shape, spiculated margins

log-likelihood image

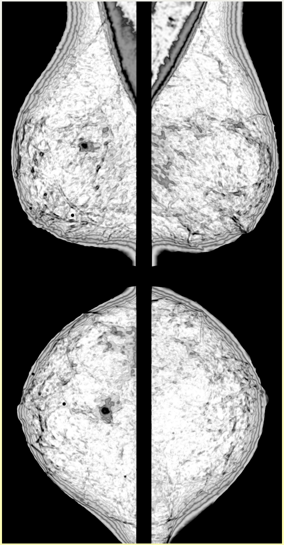


weighted log-likelihood image

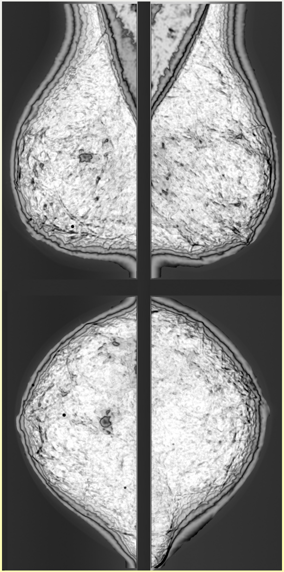


C-0143-1: mass, irregular shape, ill-defined margins

log-likelihood image



weighted log-likelihood image



Concluding Remarks

Log-Likelihood Image of Screening Mammogram:

- **purely statistical construct without any medical context**
- **aim:** to facilitate diagnostic evaluation
- **masses:** emphasized as dark regions with contour lines More
- **micro-calcifications:** dark spots of the size and form of window More
- useful for evaluation of contra-lateral findings and multi-focal lesions

GENERAL METHOD: Evaluation of Images by Local Statistical Models

- can be used to identify abnormal (suspect) locations
- need not be trained by other images (non-supervised method)
- the result of evaluation has a clear statistical interpretation
- log-likelihood image is invariant with respect to arbitrary linear transforms of the grey scale Proof
- application: fault detection, novelty detection, image forgery detection

Literatura 1/4

- Dempster A.P., Laird N.M. and Rubin D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc., B*, Vol. 39, pp.l-38.
- Grim J. (1982): On numerical evaluation of maximum - likelihood estimates for finite mixtures of distributions. *Kybernetika*, Vol.l8, No.3, pp.173-190.
- Grim J., Haindl M. (2002): A discrete mixtures colour texture model. In: Texture 2002. The 2nd International Workshop on Texture Analysis and Synthesis, Copenhagen 2002. (Chantler M. ed.). Heriot-Watt University, Glasgow 2002, pp. 59-62.
- Grim J., Haindl M. (2003): Texture Modelling by Discrete Distribution Mixtures. Computational Statistics and Data Analysis, 41, 3-4, 603-615.

Literatura 2/4

- Haindl M., Grim J., Somol P., Pudil P., Kudo M. (2004): A Gaussian mixture-based colour texture model. In: Proceedings of the 17th IAPR International Conference on Pattern Recognition. IEEE, Los Alamitos 2004, pp. 177-180.
- Haindl M., Grim J., Pudil P., Kudo M. (2005): A Hybrid BTF Model Based on Gaussian Mixtures. In: Proceedings of the 4th International Workshop on Texture Analysis and Synthesis "Texture 2005", Beijing, China, October 2005
- Grim J., Haindl M., Somol P., Pudil P. (2006): "A Subspace approach to texture modelling by using Gaussian mixtures." In: Proc. 18th IAPR International Conference on Pattern Recognition ICPR 2006. (Haralick B., Ho T. K. eds.), IEEE Computer Society, Los Alamitos 2006, pp. 235-238.

Literatura 3/4

- Grim J., Somol P., Pudil P. (2010): "Digital Image Forgery Detection by Local Statistical Models", In: Proc. 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Los Alamitos, Calif., IEEE comp. soc., eds. Echizen I. et al., pp. 579-582.
- Heath M., Bowyer K. W., Kopans D. et al. (1998): "Current State of the Digital Database for Screening Mammography." In: *Digital Mammography*, Kluwer Academic Publishers, pp. 457–460. http://marathon.csee.usf.edu/Mammography/Database.html
- McLachlan G.J. and Peel D. (2000): *Finite Mixture Models*, John Wiley & Sons, New York, Toronto.
- Schlesinger, M.I. (1968): "Relation between learning and self-learning in pattern recognition." (in Russian), *Kibernetika*, (Kiev), No. 2, 81-88.
- Titterington, D.M., Smith, A.F.M., & Makov, U.E. (1985): *Statistical analysis of finite mixture distributions.* New York: John Wiley & Sons.

Literatura 4/4

- L. Tarassenko, P. Hayton, and N. Cerneaz et al., (1995): "Novelty Detection for the Identification of Masses in Mammograms," in *4th IEEE International Conference on Artificial Neural Networks*, pp. 442–447. Cambridge, UK.
- C. J. Rose and C. J. Taylor, (2003): "A Statistical Model of Texture for Medical Image Synthesis and Analysis," in *Medical Image Understanding and Analysis*, pp. 1-4.
- C. J. Rose and C. J. Taylor, (2004): "A Generative Statis. Model of Mammographic Appearance," in *Medical Image Understanding and Analysis*, pp. 89-92.
- C. Spence, L. Parra, and P. Sajda, (2001): "Detection, Synthesis and Compression in Mammographic Image Analysis with a Hierarchical Image Probability Model," in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, L. Staib Ed.

Back

Invariance with Respect to Grey-Level Transform

Invariance Property:

log-likelihood image is invariant with respect to arbitrary linear transform of the grey scale of the original image

the transformed data and transformed mixture parameters

$$y_n = ax_n + b, \quad \tilde{\mu}_{mn} = a\mu_{mn} + b, \quad \tilde{\sigma}_{mn} = a\sigma_{mn}, \quad \mathbf{y} = T(\mathbf{x}), \ \mathbf{x} \in \mathcal{S}$$

can be shown to satisfy the EM iteration equations

$$F(\mathbf{y}|\tilde{\boldsymbol{\mu}}_m, \tilde{\boldsymbol{\sigma}}_m) = \frac{1}{a^N} F(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m), \quad \tilde{P}(\mathbf{y}) = \frac{1}{a^N} P(\mathbf{x})$$

$$q(m|\mathbf{y}) = q(m|\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}, \quad \tilde{w}_m = w_m, \quad m \in \mathcal{M}$$

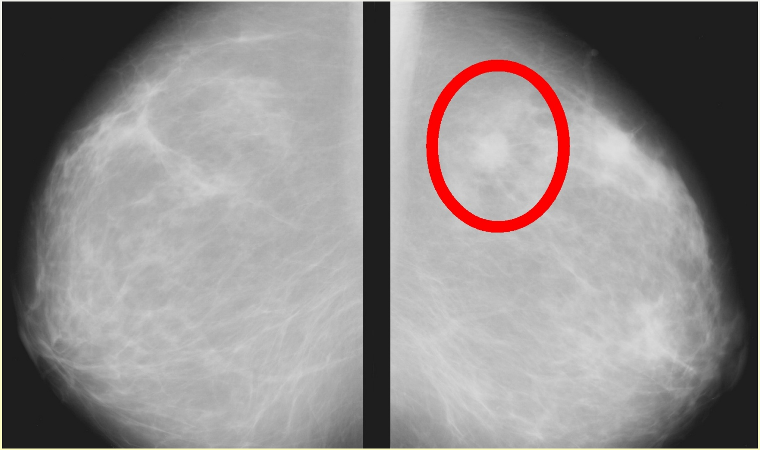
and therefore the corresponding log-likelihood values differ only by a constant

$$\log \tilde{P}(\mathbf{y}) = -N \log a + \log P(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}$$

which is removed by fixing the displayed grey-level interval Back

Identification of "Masses" by Contour-Lines

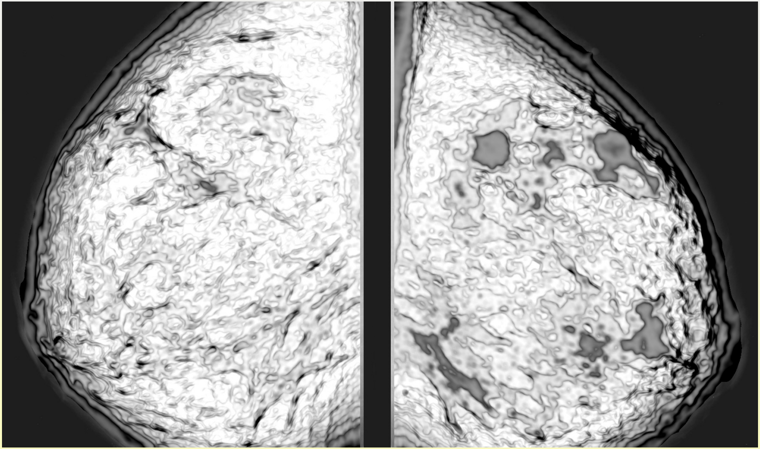
case B-3017-1: mass, lobulated-architectural distortion



Remark: The masses may be quite subtle, may have smooth boundaries and different shapes. Detection and classification of masses is more difficult than detection of micro-calcifications.

Identification of “Masses” by Contour-Lines

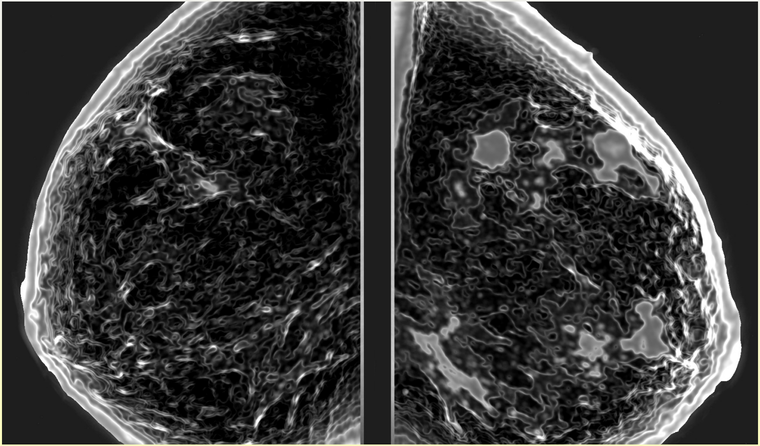
contour lines around “masses” and at the mammogram boundaries



Remark: Log-likelihood values $\log P(\mathbf{x})$ are typically dominated by a single component of the mixture. The “switching” of components at the boundaries of different areas is accompanied by decreased log-likelihood values which are visible as contour lines.

Identification of "Masses" by Contour-Lines

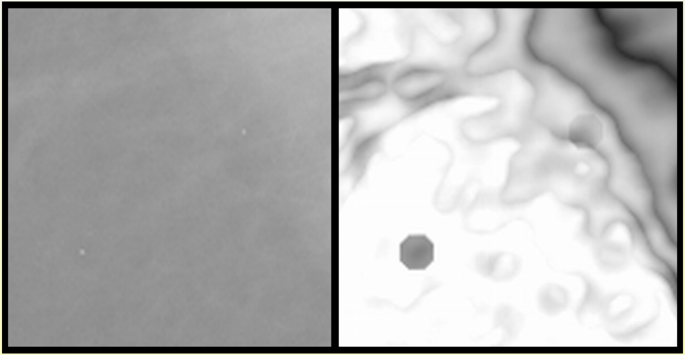
contour lines displayed by the inverse log-likelihood image



Remark: The contour lines are well visible at the mammogram boundaries characterized by continuously decreasing grey levels. The contour lines may help to evaluate possible contralateral findings or multifocal lesions.

Identification of “Micro-Calcifications” by Spots

original micro-calcifications   corresponding dark spots



Remark: Each position of the window containing a light pixel implies a lower value of $\log P(\mathbf{x})$. $\Rightarrow$ A light pixel is identified as a dark spot of window-size.

Back