Image Analysis from a Mammographic Point of View

Fredrik Georgsson
Department of Computing Science
Umeå University
Outline

Problem domain

Domain knowledge

Image Acquisition

Image Enhancement

Image Restoration

Image Segmentation

Performance Evaluation

Ground Truth

Image Enhancement

Image Restoration

Performace Evaluation

Ground Truth

Problem domain
The Problem Domain

- Breast cancer is the most common form of cancer among western women
- An efficient treatment is dependent on early diagnosis
- The best way for early diagnosis is by screening with mammography
- Efficient screening requires double reading of mammograms
- Since screening-qualified radiologists are scarce, there is a need for computer aided judging of mammograms
Mass lesions

- Smooth, well demarked surface
- Illdefined, spiculated, Irregular surface

Calcifications

- Large, coarse
- Small rounded
- Clusters of fine calcifications
- Linear branching clusters
- Irregulare shape, mixed size clusters

Secondary signs

- Skin thickening
- Ducts from lesion to nipple
- Retraction of nipple
- Retraction of skin

Parenchymal deformations

- Pulled in towards lesion
- Interrupted by lesion
- Pushed aside
Screening Round 2
April 1997 – June 1999

- Invited 28,923
- Participants 24,815 (86%)
- Selected 369 (1.5%)
- Screening mammography
- Further mammography
  - Palpation
  - Ultrasound
  - Biopsy
- Surgery 150 (0.052%)
- Cancer 110 (0.038%)
- Interval cancer
  - 60 (0.021%), missed by radiologists 7 (0.0024%)
Difficulties With Screening

• Screening mammography must be adapted to a situation where 1 out of 500 images show signs of breast cancer

• Clinical mammography can be adapted to a situation where 1 out of 10 images show signs of breast cancer
Sizes of Removed Lesions

Numbers based on the sizes of surgically removed lesions during the two screening rounds carried out in the county of Västerbotten, Sweden, 1996-1999.
• The signs of breast cancer are much more subtle in screening mammograms
• It might not even be possible to detect the actual cancer
• You might have to look for secondary signs
  – Micro-calcifications
  – Skin thickening
  – Retraction of nipple
  – Asymmetric tissue developments
Outline

Problem domain

Domain knowledge

Image Enhancement

Image Acquisition

Image Restoration

Image Segmentation

Performance Evaluation

Ground Truth

Problem domain

Image Enhancement

Image Acquisition

Image Restoration

Image Segmentation

Performance Evaluation

Ground Truth

Problem domain
Image Acquisition

- Mammograms are x-ray images of breasts
- X-rays are formed when electrons collide with a matter
- The electrons are accelerated in an electrical field
- About 99% of the energy becomes heat
Parallel projection with infinitesimal focus

Central projection with focus area
Focus 10 times larger than real focus
- The image of points are enlarged when closer to the focus
- Images of points may overlap
- Since the focus is tilted, the image of a point, depends on where the point is located
April 26, 2006  Mammography, CBA, Uppsala

Diagram showing:
- Focus
- Image of projected point
- Image plane
- Angle $\alpha$
- Lengths $d$, $e$, $a$, $b$, $c$
Typical values for the case of mammography
Outline

Problem domain

Domain knowledge

Image Acquisition

Image Restoration

Image Enhancement

Image Segmentation

Performance Evaluation

Ground Truth

Problem domain
Image enhancement

• The goal with enhancement is to improve the image for human interpretation
• This involves techniques such as
  – Histogram equalisation/specification
  – Inverting the image
  – Enlarging parts of the image
  – Using pseudo colours
Outline

- Problem domain
- Domain knowledge
- Image Acquisition
- Image Restoration
- Image Enhancement
- Image Segmentation
- Performance Evaluation
- Ground Truth
- Ground Truth

Domain knowledge

Problem domain
Image Restoration

• In order to restore an image we need a model of how the image was degraded
• If we assume that the image has been degraded via a linear shift invariant process, possibly by added noise, we have

\[ g(x,y) = f(x,y) * h(x,y) + n(x,y) \]

• Where \( g(x,y) \) is the observed image, \( f(x,y) \) is the undisturbed image, \( h(x,y) \) is the degrading and \( n(x,y) \) is the noise
Image Restoration

- $h(x,y)$ is known as the point spread function, or PSF
- Under the current assumptions $h(x,y)$ carries all information we need in order to model the degradation of the image
- Normally we gain knowledge of $h(x,y)$ by imaging objects with known properties and from the observed image we can calculate $h(x,y)$
- In fact, we often estimate the transfer function $H(u,v)$
- The transfer function and PSF are a Fourier-transform pair
Image Restoration

- A problem with x-ray imaging is that it is not a linear shift invariant process.
- This is due to the fact that the x-ray focus has an extension and that this focus-area is not parallel to the imaging plane.
- Every point in the imaged volume will be imaged as a unique area on the imaging plane.
- Thus we can conclude that the system is not shift-invariant and that it is not enough to image object at one position and base the estimate of the PSF of it.
Image Restoration

• By making assumptions regarding the
  – geometrical setting,
  – properties of the x-ray focus and,
  – imaged object
it is possible to estimate a realistic PSF.
\[ p_f(x) = \text{Focal PDF} \]
\[ p_1 = (x_1, y_1) \]

\[ \begin{align*} p_2 &= (x_2, y_2) \\
p_o^L(x) &= \text{Object PDF} \\
p_p(x) &= \text{Projected PDF} \\
p_3 &= (x_3, y_3) \end{align*} \]
Focal PDF, uniform 2mm
h = 100 cm
A = 1 cm, B = 15.5 cm, C = 30 cm
Same attenuation in all three points

[Diagram showing rays and points A, B, C]
Focal PDF, uniform 2mm
h = 100 cm
A = 1 cm, B = 15.5 cm, C = 30 cm
Same attenuation in all three points
Focal PDF, uniform 2mm
h = 100 cm
Uniform attenuation,
thickness 1 cm,
1 cm from the image plane

Focal PDF, uniform 2mm
h = 100 cm
Uniform attenuation,
thickness 29 cm,
1 cm from the image plane
Image Restoration

• Once we have an estimate of the PSF we can use a de-convolution in order to estimate the original image \( f(x,y) \) from the observed \( g(x,y) \)

• Examples of methods for de-convolution are
  – Wiener de-convolution
  – Regularized de-convolution
  – The Lucy-Richardson de-convolution
\[ g(x,y) = f(x,y) \cdot h(x,y) + n(x,y) \]
$g(x,y) = f(x,y) \cdot h(x,y) + n(x,y)$
\[ g(x,y) = f(x,y) * h(x,y) + n(x,y) \]

PSF calculated from mammographic setting with uniform attenuation and tissue thickness of 9 cm. Pixel size 44\( \mu m \).
Image Restoration

• In some cases the degradation of the image is better modelled as

\[ g(x,y) = f(x,y) + h(x,y) + n(x,y) \]
Image Restoration

- In order to reduce the effect of secondary radiation in x-ray images we introduce a lead screen
X-ray of orthopaedic marker
diameter approx 1mm

Log magnitude of Fourier spectra

April 26, 2006
Mammography, CBA, Uppsala
Outline

Problem domain

Image Acquisition

Image Restoration

Image Enhancement

Image Segmentation

Performance Evaluation

Ground Truth

Domain knowledge

April 26, 2006
Image Segmentation

• In order to analyse a mammogram it has to be segmented into its constituent parts
  – The breast area
  – The pectoralis muscle
  – The mamilla
  – The glandular disc
  – Fat (compressed and uncompressed)
  – ....
Optimal: 0.3608
Mean: 0.3705
Based on 2000 tests
Breast Border

- Based on the found threshold the image is segmented into breast and background.
- Using the original intensities found in the background a polynomial surface is constructed and subtracted from the image.
- The procedure is repeated until the changes are small.
Breast Border

April 26, 2006

Mammography, CBA, Uppsala
Texture Classification

Texture descriptions

Mammogram

Texture representation

ANN

Compressed fat
Uncompressed fat
Glandular tissue
Muscular tissue
Texture Representations

- Hard, fixed sized neighbourhood (15 x 15)
- Representation
  - Runlength matrices
  - Fourier Spectra
- Descriptions
  - RA-radon (direction)
  - Q2-2nd percentile (intensity)
  - RS-relative smoothness (roughness)
  - S-entropy (roughness)
  - RLM4-run length (roughness)
  - $\beta$ -signal ratio power (roughness)
Outline

- Problem domain
- Domain knowledge
- Image Acquisition
- Image Restoration
- Image Enhancement
- Image Segmentation
- Performance Evaluation
- Ground Truth

Ground Truth
Problem domain
Ground Truth Assessment

• In order to evaluate the performance of an image segmentation we need a golden truth to compare to

• Knowledge of this **ground truth** can be found by
  – Using simulated data
  – Establishing it by other examinations (for instance biopsy)
  – Letting domain experts state it
Ground Truth Assessment

• Since we have variations it is non-trivial to combine ground truths assessed by domain experts

• One domain expert might assess the ground truth differently if asked several times (intra variance)

• Different domain experts might assess the ground truth differently (inter variance)
Ground Truth Assessment

- Suppose we have $K$ experts that have all marked a feature $A$
- One can define $\Lambda$ to be a measure of agreement

\[
\Lambda^p_i = \frac{\mu(A_1^p \cap A_2^p \cap \ldots \cap A_K^p)}{\mu(A_1^p \cup A_2^p \cup \ldots \cup A_K^p)},
\]

If we have any variance (inter or intra), the $\Lambda$-measure of agreement is of no use.
Performance Evaluation
Performance Evaluation

Based on all images in case except r0082
Performance Evaluation

Based on all images in case except A0082
Performance Evaluation

![Performance Evaluation Diagram]

April 26, 2006
Mammography, CBA, Uppsala
Performance Evaluation

April 26, 2006

Mammography, CBA, Uppsala
Performance Evaluation

C0047
-0.766

F0047
0.304

r0047
0.77

B0047
-0.907

A0047
0.177

D0047
0.422
**Result**

\[
\text{RankSum}_i = \sum_j E_{ij} * j,
\]

where \( i = 1, \ldots, 6 \), \( j = 1, \ldots, 6 \) and \( E_{ij} \) is a histogram over the rank value of the markings.

<table>
<thead>
<tr>
<th>Expert</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>( A_5 )</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>( A_6 )</td>
<td>13</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>12</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>( A_5 )</td>
<td>16</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( A_6 )</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>23</td>
</tr>
</tbody>
</table>

---

April 26, 2006  
Mammography, CBA, Uppsala
### Result

#### Table 1

<table>
<thead>
<tr>
<th>Rank</th>
<th>Sum</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>115</td>
<td>$A_2$</td>
</tr>
<tr>
<td>2</td>
<td>137</td>
<td>$A_4$</td>
</tr>
<tr>
<td>3</td>
<td>139</td>
<td>$A_6$</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
<td>$A_5$</td>
</tr>
<tr>
<td>5</td>
<td>154</td>
<td>$A_3$</td>
</tr>
<tr>
<td>6</td>
<td>155</td>
<td>$A_1$</td>
</tr>
</tbody>
</table>

#### Table 2

<table>
<thead>
<tr>
<th>Rank</th>
<th>Sum</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>104</td>
<td>$A_5$</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>$A_2$</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>$A_4$</td>
</tr>
<tr>
<td>4</td>
<td>138</td>
<td>$A_3$</td>
</tr>
<tr>
<td>5</td>
<td>148</td>
<td>$A_1$</td>
</tr>
<tr>
<td>6</td>
<td>213</td>
<td>$A_6$</td>
</tr>
</tbody>
</table>
Conclusions

• It is important to
  – have an understanding of the problem area
  – roughly speak the language of the domain experts
  – have a large toolbox of image analysis methods at once disposal
  – not underestimate the difficulties involved
Thanks for your attention