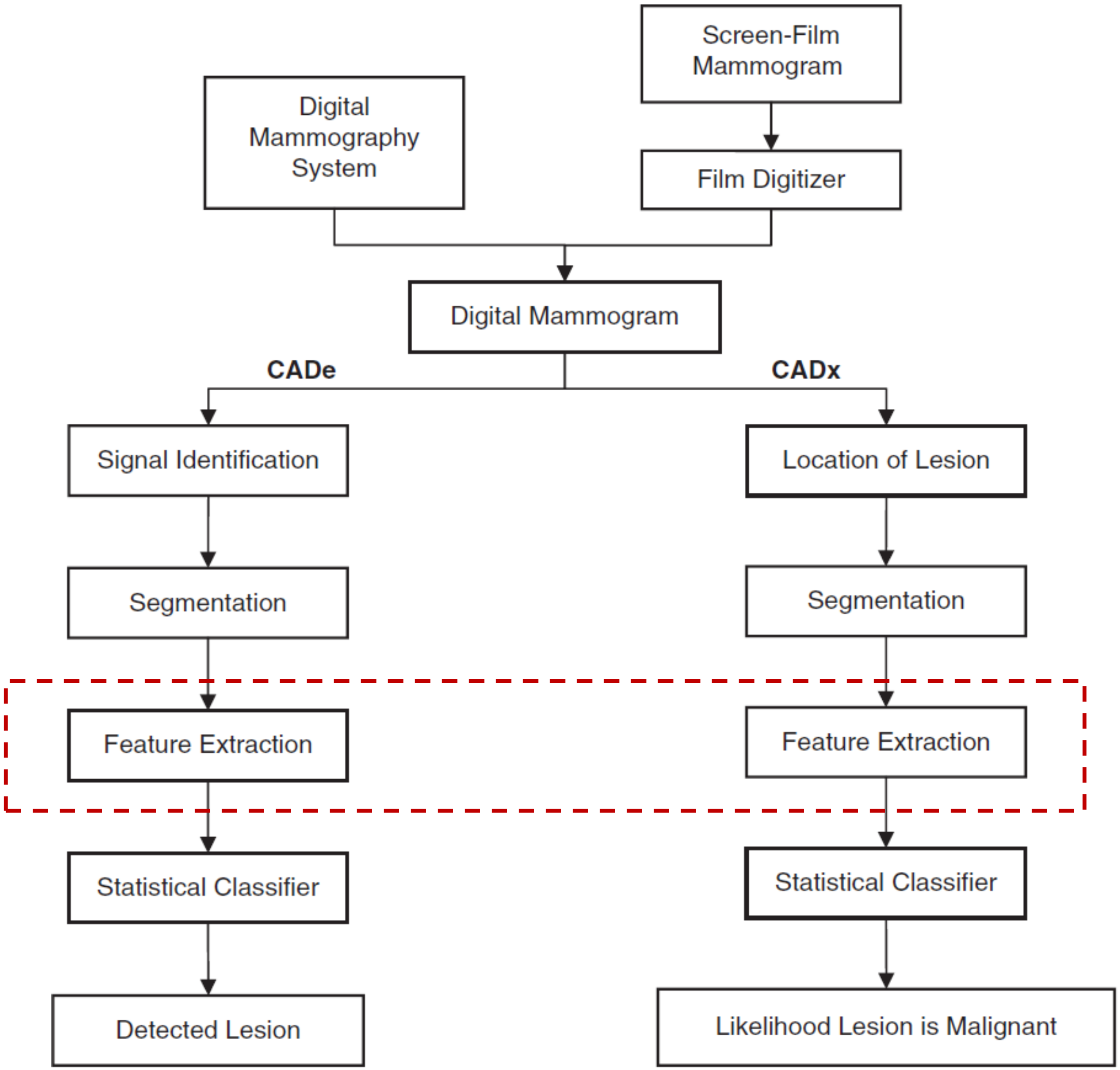
ADVANCED TOPICS IN BIOMEDICAL ENGINEERING

Topic 4: Feature Extraction

Prof. Yasser Mostafa Kadah

CAD

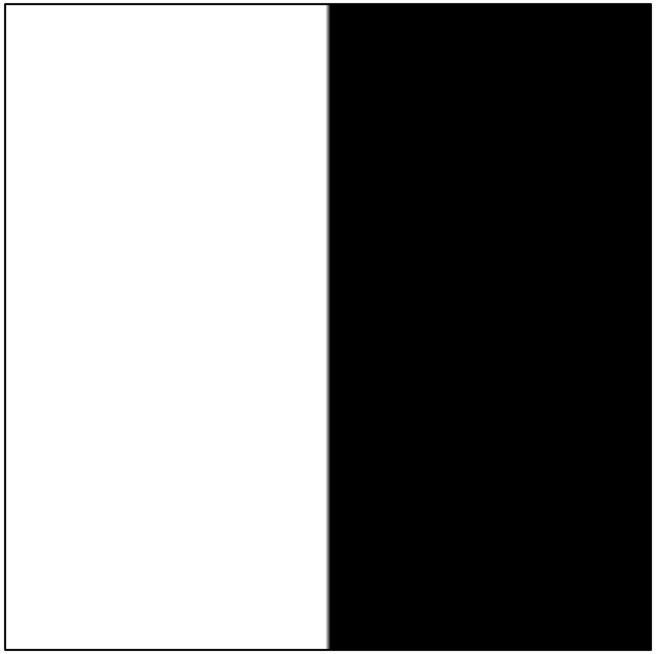

Screen-Film Mammogram
Digital Mammography System
Film Digitizer
Digital Mammogram
CADe
CADx
Signal Identification
Location of Lesion
Segmentation
Segmentation
Feature Extraction
Feature Extraction
Statistical Classifier
Statistical Classifier
Detected Lesion
Likelihood Lesion is Malignant

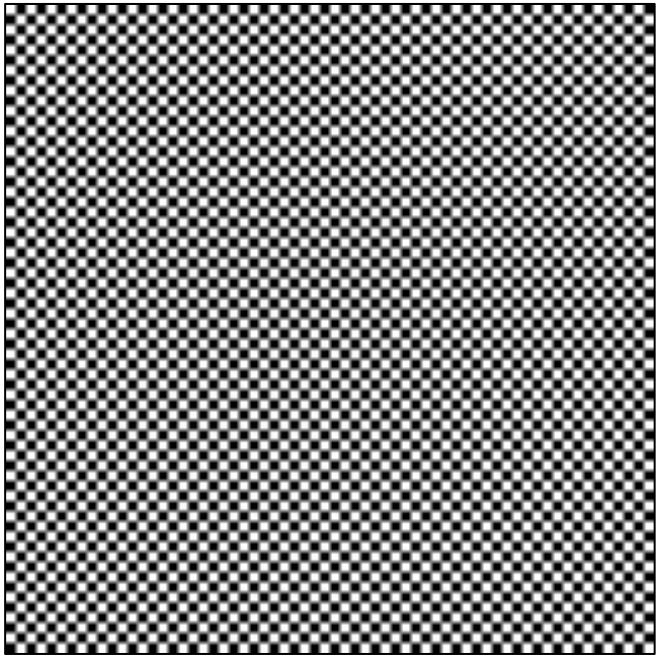
Feature Categories

- Features are quantitative measures of texture that describe salient characteristics in the image
- Spatial domain features
  - First order statistical or Histogram-based features
  - Higher order statistical features
- Transform domain features
  - Fourier descriptors
  - Wavelet features

First Order Statistical Features

- Depend only on pixel values and independent of spatial distribution of pixels
  - Example: images below have same first order features





First Order Statistical Features

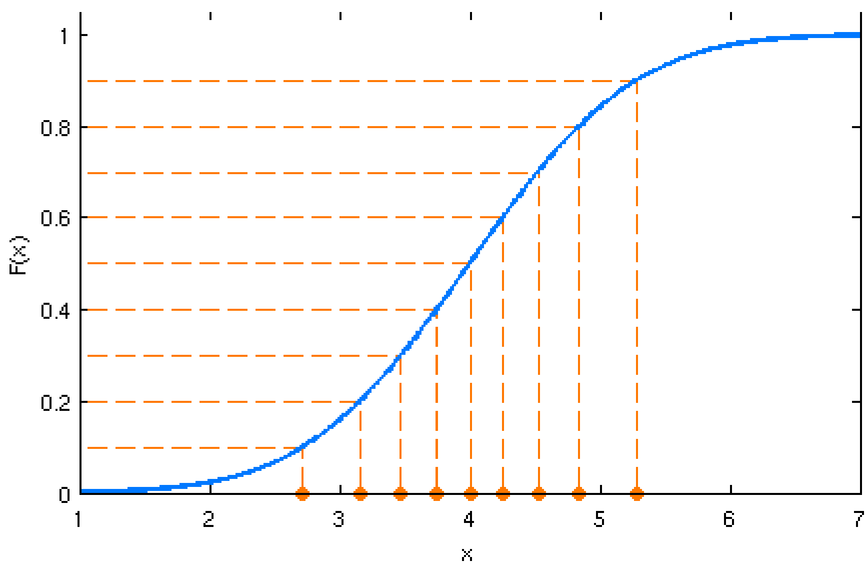
- Example features:

Moment	Expression	Measure of texture
Mean	$m = \sum_{i=0}^{L-1} z_i \; p(z_i)$	A measure of average intensity
Standard deviation	$\sigma = \sqrt{\mu_2 \; (z)} = \sqrt{\sigma^2}$	A measure of average contrast
Smoothness	$R = 1 - 1/(1 + \sigma^2)$	Measures the relative smoothness of the intensity in a region.
Third Moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \; p(z_i)$	Measures the skew ness of a histogram
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$	Measures the uniformity. of intensity in the histogram
Entropy	$e = -\sum_{i=0}^{L-1} p(z_i) \; log_2 \; p(z_i)$	A measure of randomness

First Order Statistical Features

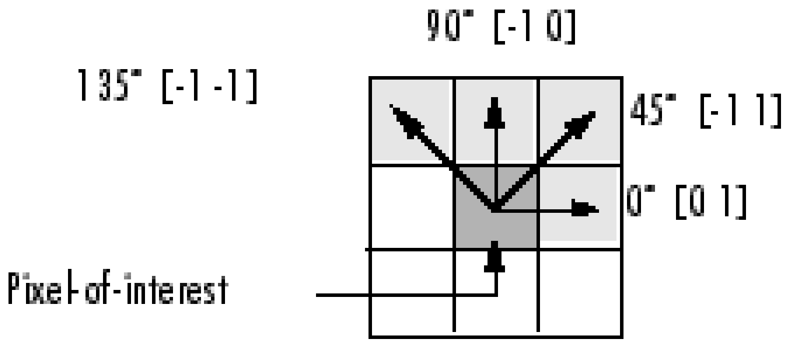
- Quantiles (or percentiles) are points taken at regular intervals from the cumulative distribution function (CDF) of a random variable
  - Dividing ordered data into essentially equal-sized data subsets is the motivation for **q-quantiles**; quantiles are the data values marking the boundaries between consecutive subsets
  - Common to use 0.1, 0.2, …, 0.9 as features
  - 0.5-quantile is the <u>median</u>

Matlab function "quantile(data,p)":
Returns quantile of the values in "data" for the cumulative probability or probabilities "p" in the interval [0,1]



Higher Order Statistical Features

- Depend on pixel values and their spatial inter-relationships.
- Example: Cooccurrence matrix (GLCM) features
  - GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image
  - GLCM is constructed by observing pairs of image cells distance *d* from each other and incrementing the matrix position corresponding to the grey level of both cells
  - Different realization depending on distance and angle



Matlab function "graycomatrix": Create gray-level co-occurrence matrix from image

Higher Order Statistical Features

1	1	7	5	3	2
5	1	6	1	2	5
2	0	6	4	1	2
4	3	4	5	5	1
0	7	1	7	6	2
7	4	6	2	6	2

Image f

	1	2	3	4	5	6	7	8
1	1	2	0	0	0	1	1	0
2	0	0	0	0	1	1	0	0
3	0	1	0	1	0	0	0	0
4	0	0	1	0	1	0	0	0
5	2	0	1	0	1	0	0	0
6	1	3	0	0	0	0	0	1
7	0	0	0	0	1	1	0	2
8	1	0	0	0	0	2	2	1

Co-occurrence matrix C

Higher Order Statistical Features

Descriptor	Explanation	Formula
Maximum probability	Measures the strongest response of C. The range of values is [0,1]	$$\max_{(i,j)}(P_{ij}) \qquad (2.5)$$
Correlation	A measure of how correlated a pixel is to its neighbour over the entire image. Range of values is 1 to -1, corresponding to perfect positive and perfect negative correlations. This measure is not defined if either standard deviation is zero.	$$\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(i-m_r)(j-m_c)P_{ij}}{\sigma_r\sigma_c}$$ $$\sigma_r \neq 0; \sigma_c \neq 0 \qquad (2.6)$$
Contrast	A measure of intensity contrast between a pixel and its neighbour over the entire image. The range of values is 0 (when C is constant) to $(k-1)^2$	$$\sum_{i=1}^{k}\sum_{j=1}^{k}(i-j)^2 P_{ij} \qquad (2.7)$$

Higher Order Statistical Features

Uniformity (also called Energy)	A measure of uniformity in the range [0,1]. Uniformity is 1 for a contrast image.	$$\sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}^2 \quad (2.8)$$
Homogeneity	Measures the spatial closeness of the distribution of elements in C to the diagonal. The range of values is [0,1], with the maximum being achieved when C is a diagonal matrix.	$$\sum_{i=1}^{k}\sum_{i=1}^{k} \frac{P_{ij}}{1+\lvert i-j\rvert} \quad (2.9)$$
Entropy	Measures the randomness of the elements of C. The entropy is 0 when all P_{ij}'s are 0 and is maximum when all P_{ij}'s are equal.	$$-\sum_{i=1}^{k}\sum_{i=1}^{k} P_{ij}\log_2 P_{ij} \quad (2.10)$$

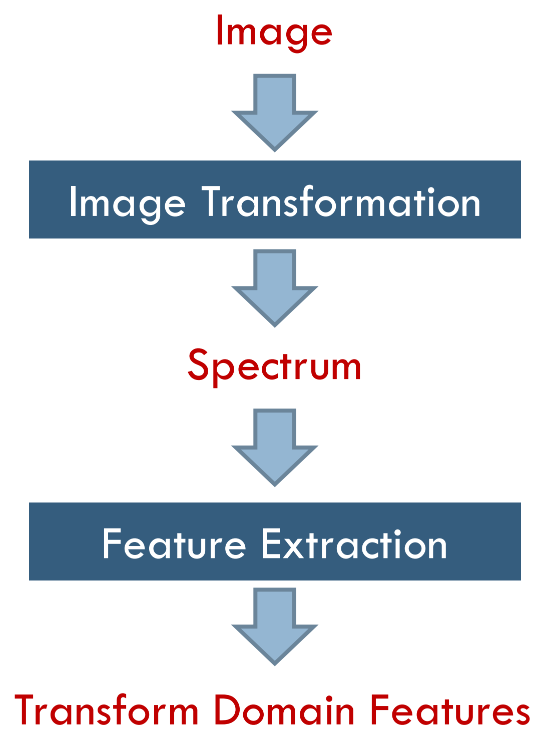
Higher Order Statistical Features

- Matlab "graycoprops"
  - Properties of gray-level co-occurrence matrix

Property	Description	Formula
`'Contrast'`	Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image. `Range = [0 (size(GLCM,1)-1)^2]` Contrast is 0 for a constant image.	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
`'Correlation'`	Returns a measure of how correlated a pixel is to its neighbor over the whole image. `Range = [-1 1]` Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is `NaN` for a constant image.	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$
`'Energy'`	Returns the sum of squared elements in the GLCM. `Range = [0 1]` Energy is `1` for a constant image.	$\sum_{i,j} p(i,j)^2$
`'Homogeneity'`	Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. `Range = [0 1]` Homogeneity is 1 for a diagonal GLCM.	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$

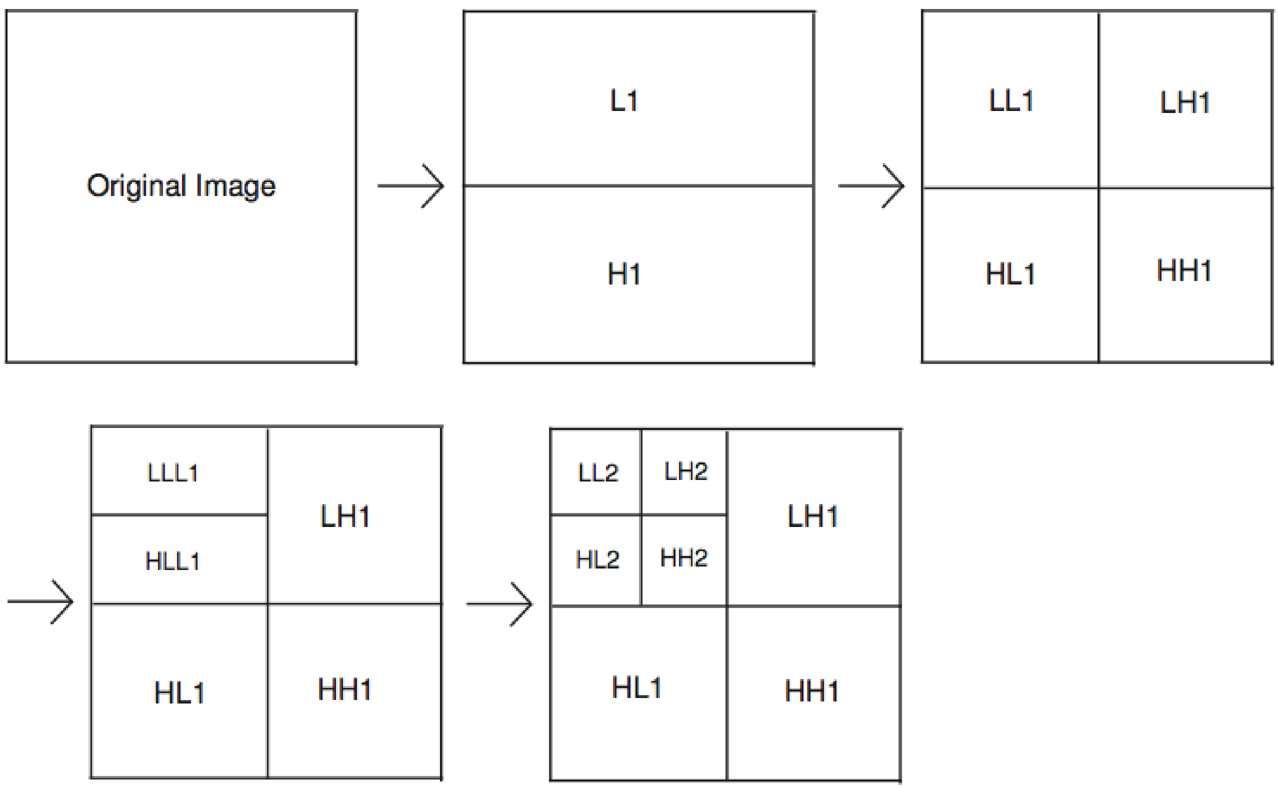
Transform Domain Features

- Features computed AFTER transformation to another domain
  - Discrete Fourier transform
  - Discrete cosine transform
  - Discrete wavelet transform



Transform Domain Features

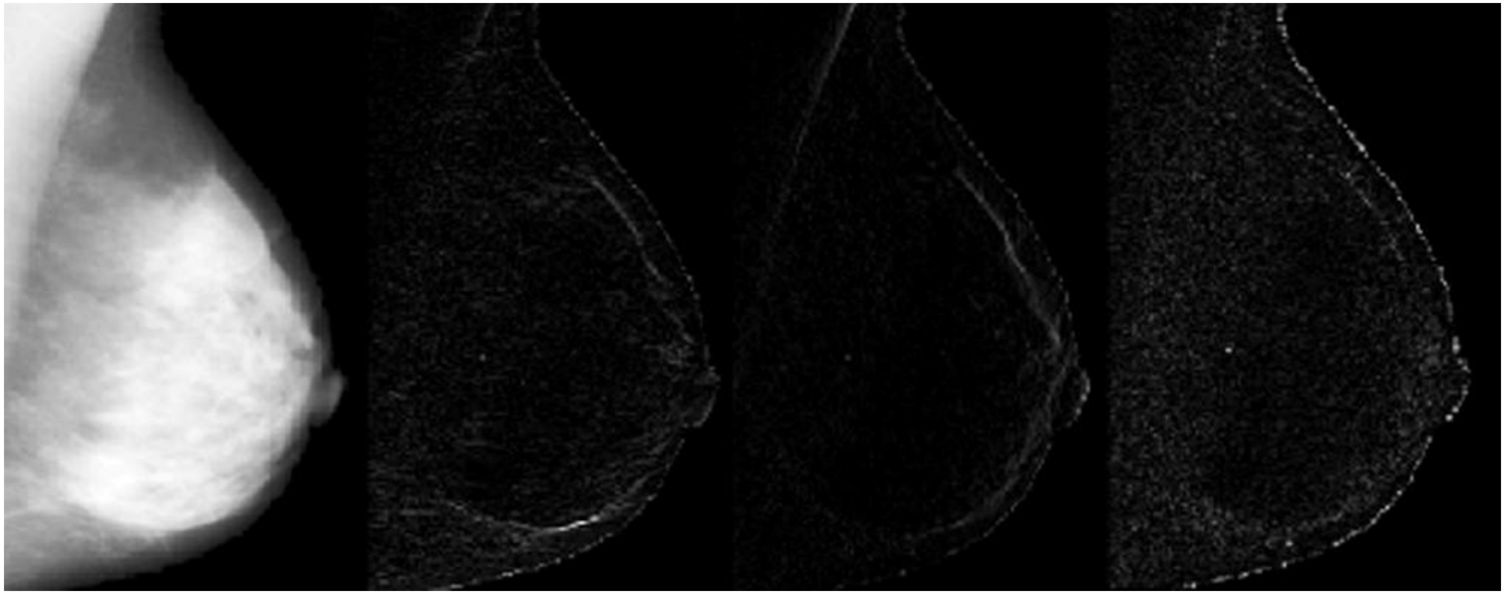
- Example: Wavelet decomposition
  - Selection of basic wavelet type and size and number of levels



(c)

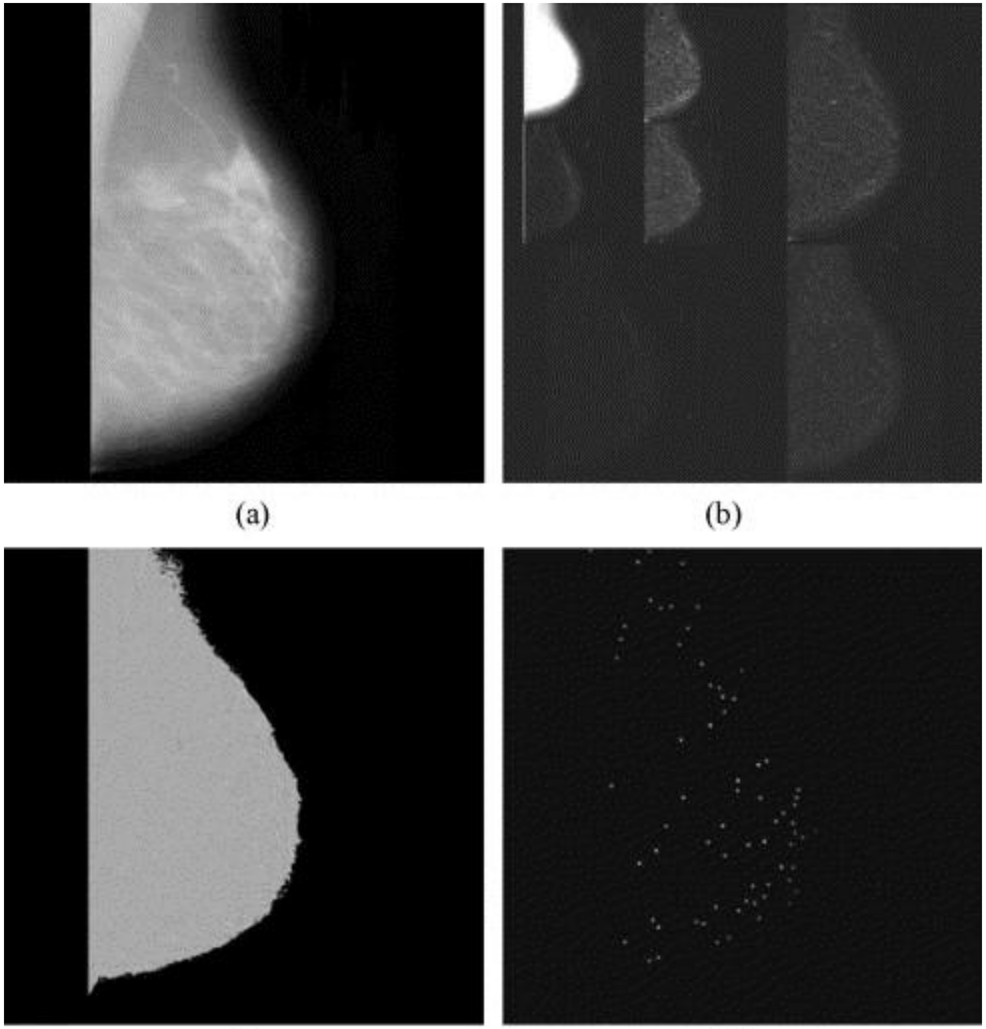
Transform Domain Features

- Wavelet decomposition of a mammogram
  - Different details in different scales
  - Consider each as a spatial image and proceed with spatial feature extraction



Transform Domain Features

- Example: Microcalcification detection in Wavelet domain


(a) (b) (c) (d)

Sung-Nien Yu, Kuan-Yuei Li, Yu-Kun Huang, "Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model," *Computerized Medical Imaging and Graphics*, vol. 30, no. 3, pp. 163–173, April 2006.

Final Feature Extraction Notes

- Newer features based on image modeling
  - Markov random model
  - ARMA model
  - Fractal models
- Next step: feature selection
  - Not all features are correlated with disease
  - Must include only relevant features to avoid misclassification
  - Normalization of feature values is necessary preprocessing step before training calssifiers

Assignments

- Read the thesis titled "Texture Descriptors applied to Digital Mammography" (Google search to download it)
- Start with the miniMIAS database and select 10 images with masses and 10 with normal texture.
  - Consider ROI size of 32x32
  - Create an array of 30 ROIs for normal and abnormal pathologies (statistical unit: lesion NOT patient)
  - Compute at least 20 features for each ROI