# Digital Image Processing

14 December 2006

**Dr. ir. Aleksandra Pizurica**
Prof. Dr. Ir. Wilfried Philips

Aleksandra.Pizurica @telin.UGent.be
Tel: 09/264.3415

UNIVERSITEIT GENT
**Telecommunicatie en Informatieverwerking**

INSTITUUT VOOR PERMANENTE VORMING




## Filters adapted to local noise statisics

Assume: $g(x,y) = f(x,y) + n(x,y)$

The estimate $\hat{f}$ that minimizes the mean squared error $E\{(f-\hat{f})^2\}$ is

$$\hat{f}(x,y) = g(x,y) - \frac{\sigma_n^2}{\sigma_f^2 + \sigma_n^2}(g(x,y) - \bar{f}(x,y))$$

Where $\bar{f}$ and $\sigma_f$ denote the mean and the standard deviation of $f$, resp.

We have $\sigma_f^2 + \sigma_n^2 = \sigma_g^2$ and if $n$ is **zero mean** process: $\bar{f}(x,y) = \bar{g}(x,y)$

What we need to know to proceed?

(1) Assuming $f$ (and $g$) is ergodic process: $\bar{g}(x,y) = m_L$ , $\sigma_g^2 = V_L$

local mean local variance

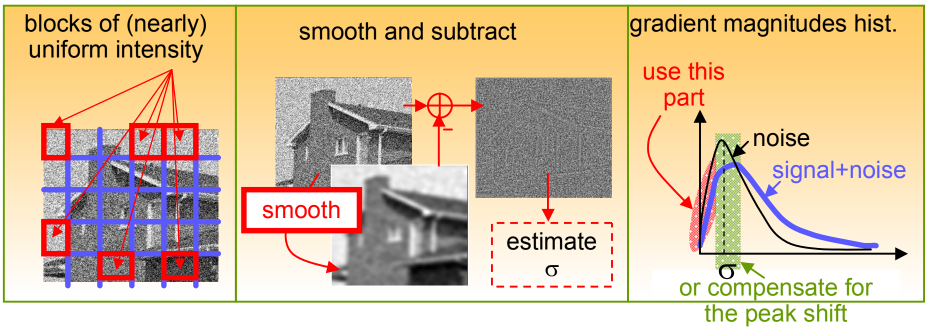
(2) We need an estimate of the noise variance $\hat{\sigma}_n^2$

$$\hat{f}(x,y) = g(x,y) - \frac{\hat{\sigma}_n^2}{V_L}(g(x,y) - m_L))$$

# Estimation of noise parameters

Approaches for estimating standard deviation σ of AWGN noise:

- block-based: use regions (blocks) with the least signal variation
- smoothing-based: estimate σ from the difference between the image and its smoothed version, assuming that this difference is pure noise
- gradient-distribution based: if image is pure noise distribution of gradient magnitudes is Rayleigh peaked at σ







© A. Pizurica, Universiteit Gent, 2006

# Suppression of multiplicative noise

Two usual approaches:

Input image $I$ → $\log(I+1)$ → Filter for additive noise → $J$ → $\exp(J)\text{-}1$ →

The logarithmic operation transforms multiplicative noise into additive

Input image $I$ → linearize → Filter for additive noise → $J$

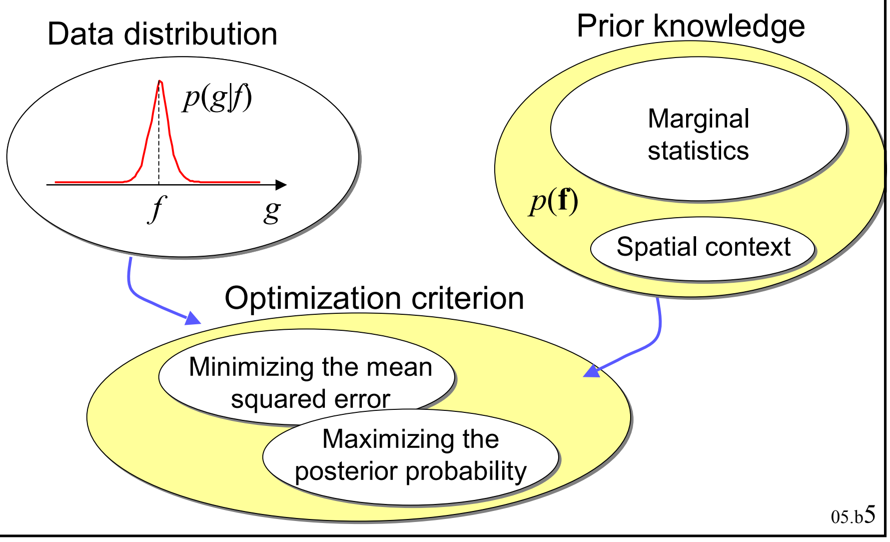
A second approach is based on “linearizing” the multiplicative noise model, e.g., by uisng the first-order Taylor series expansion

This is the basis of the well-known Lee filter for speckle noise

# Bayesian Approach

Bayesian approach using prior knowledge about the unknown image

Denote $g=f+n$









# Bayesian approach

Denote

$g$ – observed noisy data;

$f$ – ideal noise-free data:

Maximum a posteriori (MAP) estimate:

$$\hat{f}_{MAP} = \arg\max_{f \in \Im} p(f \mid g) \;= \arg\max_{\mathbf{f} \in \Im} \frac{p(g \mid f)P(f)}{P(g)}$$

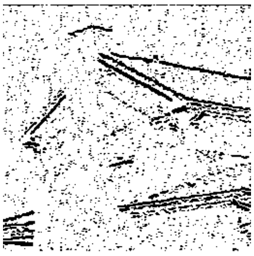
$$= \arg\max_{\mathbf{f} \in \Im} p(g \mid f)P(f)$$

Minimum mean squared error (MMSE) estimate

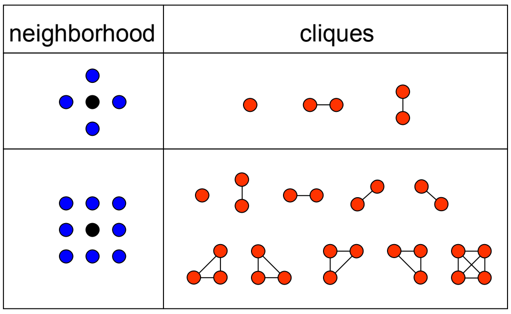
$$\hat{f}_{MMSE} = E(f \mid g) = \int_{-\infty}^{\infty} f\, p(f \mid g) df \;= \frac{\int_{-\infty}^{\infty} f p(g \mid f) p(f) df}{\int_{-\infty}^{\infty} p(g \mid f) p(f) df}$$

If $p(f)$ is Gaussian and noise is also additive Gaussian (then $p(g|f)$ is Gaussian) the MMSE estimate reduces to the Wiener filter

## Markov Random Field (MRF) prior model



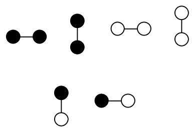
$\mathbf{f}=\{f_1,\ldots,f_n\}$




$$P(\mathbf{f}) = \frac{1}{Z}\exp\left\{-\sum_{C\in\varsigma} V_C(\mathbf{f})\right\}$$

clique potentials

Example: penalize isolated peaks



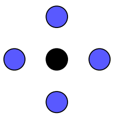
negative potential

positive potential





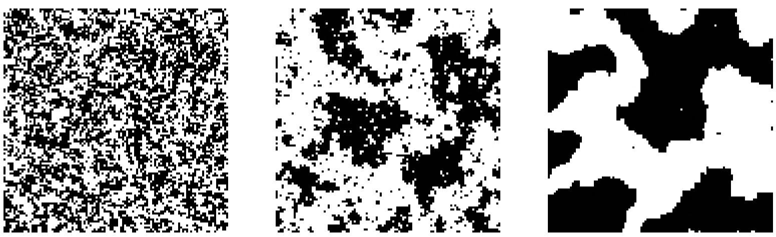

## Ising model: an MRF model for binary images



Pair-site clique potential

$$V_2(f_l, f_k) = \begin{cases} -\gamma, & f_l = f_k \\ \gamma, & f_l \neq f_k \end{cases}, \gamma > 0$$

Samples form the Ising model

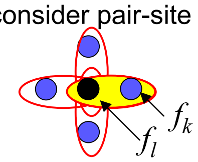


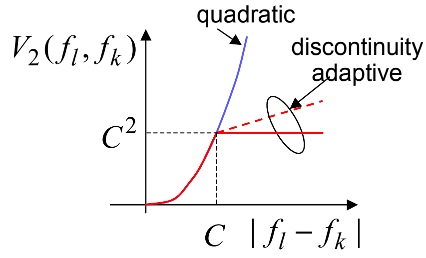
$\longrightarrow \gamma$

Spatial position $(x,l)$ is denoted by a single index (raster scanning)

# MRF models for grayscale images

We consider pair-site cliques only




- Quadratic (auto-normal) model:

$$V_2(f_l, f_k) = \alpha (f_l - f_k)^2$$

→ This model tends to oversmooth the image (destroys edges).

- An example of a discontinuity adaptive model:

$$V_2(f_l, f_k) = \begin{cases} \alpha (f_l - f_k)^2, & \text{if } |f_l - f_k| < C \\ const\,(= \alpha C^2), & \text{otherwise} \end{cases}$$

→ Differences that are bigger than a threshold are less penalized because these are likely to originate from actual edges






# Image denoising using MRF priors…

Assume: $g(x,y) = f(x,y) + n(x,y)$

MAP estimate: $\hat{\mathbf{f}} = \arg\max_{\mathbf{f}\in\Im} p(\mathbf{f}\,|\,\mathbf{g}) \;= \arg\max_{\mathbf{f}\in\Im} \underbrace{p(\mathbf{g}\,|\,\mathbf{f})}_{\text{data distr.}} \underbrace{P(\mathbf{f})}_{\text{prior}}$

For white Gaussian noise with zero mean and standard deviation σ:

$$p(\mathbf{g}\,|\,\mathbf{f}) = A \exp\Big(-\sum_l (g_l - f_l)^2 / 2\sigma^2\Big)$$

MRF prior: $p(\mathbf{f}) = Z \exp\Big(-\sum_{c\in\zeta} V_C(\mathbf{f})\Big) = Z \exp\Big(-\sum_{\langle k,l\rangle} V_2(f_k, f_l)\Big)$ ← over all pairs of pixels

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}\in\Im} \log\big(p(\mathbf{g}\,|\,\mathbf{f})P(\mathbf{f})\big) \;= \arg\min_{\mathbf{f}\in\Im} \underbrace{E(\mathbf{f}\,|\,\mathbf{g})}_{\text{posterior energy}}$$

$$E(\mathbf{f}\,|\,\mathbf{g}) = \underbrace{\sum_l (g_l - f_l)^2}_{\text{data fit term}} + \underbrace{\lambda \sum_{\langle k,l\rangle} V_2(f_k, f_l)}_{\text{prior knowledge (context)}}$$

(λ is a constant)

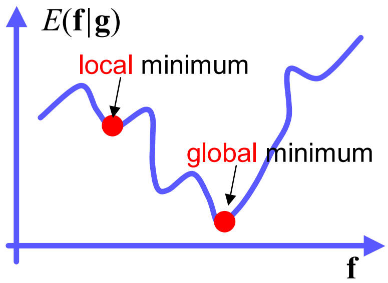
# …Image denoising using MRF priors

Assume: $g(x,y) = f(x,y) + n(x,y)$

MAP solution: minimize the posterior energy $E(\mathbf{f}|\mathbf{g})$

$$E(\mathbf{f}\,|\,\mathbf{g}) = \sum_l (g_l - f_l)^2 + \lambda \sum_{\langle k,l \rangle} V_2(f_k, f_l)$$

**data fit term** | **prior knowledge** (context)




Greedy algorithms tend to end up in a local minimum!

Solutions:

- Random search algorithms – allow occasional increase in posterior energy in order to achieve the gloobal minimum (example: Metroplois sampler)
- Genetic algoritms



version: 14/12/2006
© A. Pizurica, Universiteit Gent, 2006

# Total variation (TV) image denoising

Reasoning similar as in the Bayesian approach with MRF priors →

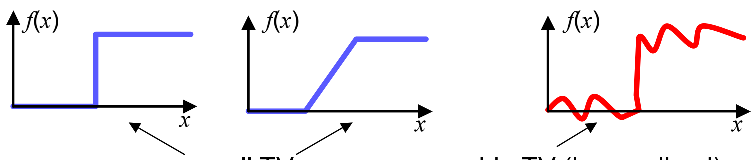
→ Find the ideal image f that minimizes the total variation energy $E_{tv}$:

$$E_{tv} = \iint \left( (f(x,y) - g(x,y))^2 + \lambda |\nabla f(x,y)| \right) dxdy$$

**data fit term** | **regularization** term (**variation**)

The assumed underlying image model is in this case a "cartoon model"

- Images consist of big uniform areas with $\nabla f(x,y) \approx 0$, separated by edges having big $\nabla f(x,y)$
- Images with noise have throughout $|\nabla f(x,y)|$ bigger than ideal noise-free image



# Total variation (TV) : a practical method

For white Gaussian noise with standard deviation $\sigma_n$, the regularization paramreter $\lambda$ is sometimes calculated such that

$$\frac{1}{\Omega}\underbrace{\iint (\hat{f}(x,y)-g(x,y))^2 dxdy}_{\text{Measured squared error between the estimated ideal image and the noisy image}} = \underbrace{\sigma_n^2}_{\text{The expected square error between noise-free image and the measured image}}$$

$\Omega$ the set of all image pixels

A practical method:

- start with $\lambda=\lambda_0$ and compute $\hat{f}(x,y)$ by minimizing $E_{tv}$
- Increase or decrease $\lambda$ depending on whether the measured squared error is to high or too smaal and recompute $\hat{f}(x,y)$
- Repeat until the deviation is sufficiently small

Slow procedure!






# Total-variation: example

Original

Image with white noise $\sigma$=20

Result with optimal $\lambda$

Regularization penalizes not only noise, but also texture

$\Rightarrow$ Grass is blurred

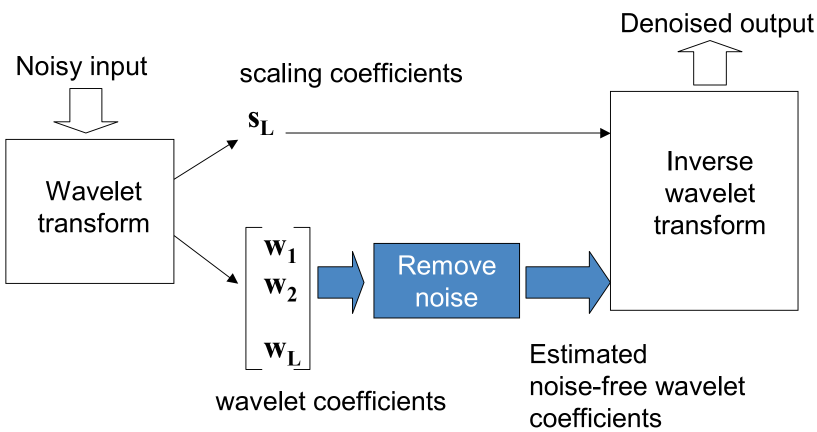
Applications:

- Noise reduction in images without texture
- Pre-processing for segmentation

# Wavelet domain denoising




## Wavelet domain denosing



# Noise model and notation

image domain: $v_l = f_l + \vartheta_l;$ $\vartheta_l \sim N(0, \sigma_{in})$

wavelet domain: $w_{j,l}^{d} = y_{j,l}^{d} + n_{j,l}^{d};$ $n_{j,l}^{d} \sim N(0, \sigma_j^d)$

orientation

resolution level

spatial position

$$\left(\sigma_j^d\right)^2 = S_j^d \sigma_{in}^2$$

$$S_j^{LH,HL} = \left(\sum_k g_k^2\right)\left(\sum_k h_k^2\right)^{2j-1} \qquad S_j^{HH} = \left(\sum_k g_k^2\right)^2\left(\sum_k h_k^2\right)^{2(j-1)}$$

$g_k$ and $h_k$ are the coefficients of the high-pass and low-pass decomposition filters

**Orthogonal** wavelet transform maps white noise into white noise and $\sigma_j^d = \sigma = \sigma_{in}$ for all $j,d$






# Noise variance estimation

- Often the value of the input noise is unknown
- Noise has to be estimated from the observed noisy signal eliminating the influence of the actual signal
- A median measurement is highly insensitive to outliers
- Meadian Absolute Deviation (MAD) estimator

$$\hat{\sigma} = Median(|\, \mathbf{w}_1^{HH} \,|)/0.6745$$

# Ideal coefficient selection and attenuation

**Noise model** $w = y + n$; $n \sim N(0,\sigma)$

**Consider a nonlinear estimate** $\hat{y} = \theta(y)w$

**What is ideal $\theta(y)$ in the MSE sense?**

One can show that $E\{(y - w\theta(y))^2\} = y^2(1-\theta(y))^2 + \sigma^2\theta(y)^2$

**The estimator $\theta_{ideal}(y)$ that minimizes this error is** ideal attenuator

$$\theta_{ideal}(y) = \frac{y^2}{y^2 + \sigma^2}$$

**If we restrict** $\theta(y) \in \{0,1\}$ this becomes

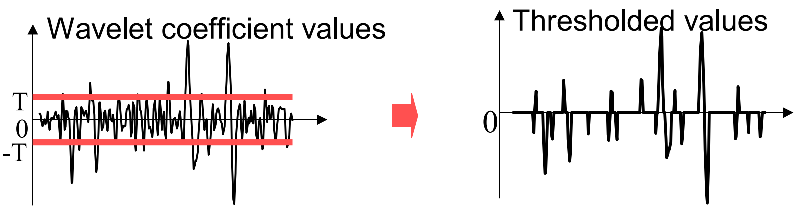
$$\theta_{ideal/sel.}(y) = \begin{cases} 1, & \text{if } |y| > \sigma \\ 0, & \text{if } |y| \le \sigma \end{cases}$$

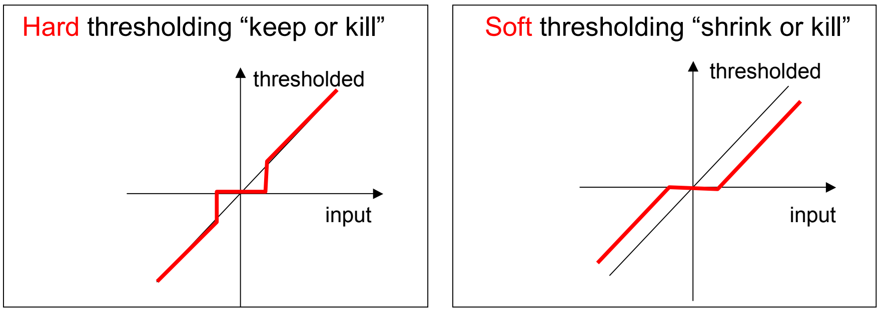
**oracle thresholding**






# Denoising by wavelet thresholding



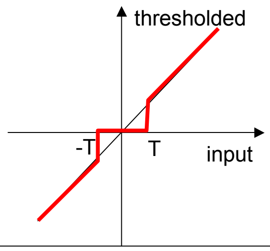



# Denoising by wavelet thresholding

$w = y + n$    $w$ – noisy coefficient;
$y$ – noise-free coefficient; $n$ – i.i.d. Gaussian noise

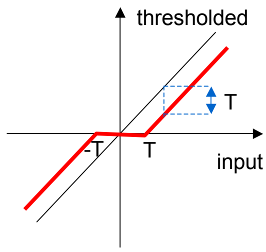
$$\hat{y}_{ht} = \begin{cases} 0, & |w| < T \\ w, & |w| \geq T \end{cases}$$

Hard thresholding “keep or kill”




$$\hat{y}_{st} = \begin{cases} 0, & |w| < T \\ \operatorname{sgn}(w)(|w| - T), & |w| \geq T \end{cases}$$

Soft thresholding “shrink or kill”





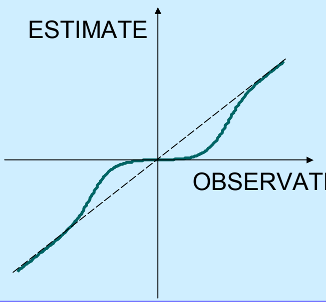




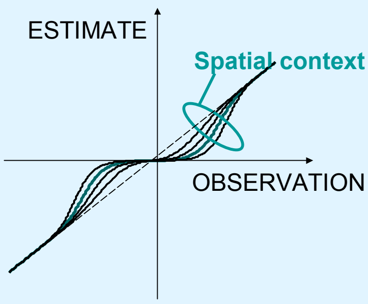
# Wavelet shrinkage

- Hard- and soft-thresholding are examples of wavelet shrinkage
- Shrinkage can result from Bayesian methods too

Subband-adaptive shrinkage

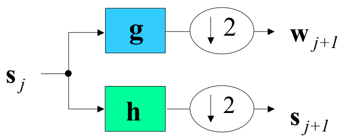



Spatially adaptive shrinkage



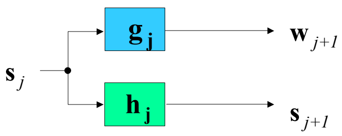
# Non-decimated wavelet transform

(bi-) orthogonal DWT




not shift invariant!

Non-decimated WT




Algorithm *à trous*

$\mathbf{g}_\mathbf{j}$ **:** $2^{j-1}$ zeros inserted between each two coefficients of $\mathbf{g}$

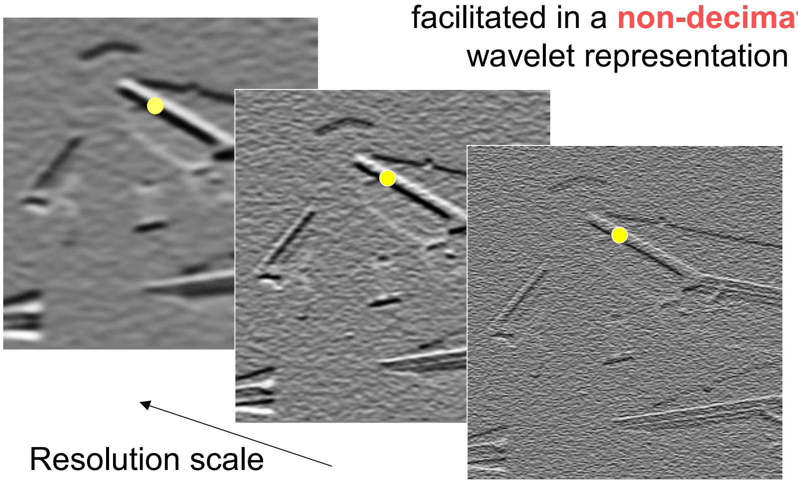
Also called *overcomplete* and *stationary* wavelet transform





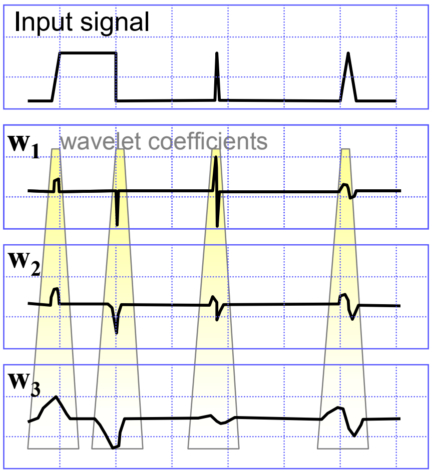

# Non-decimated Representation

Inter-scale comparisons are facilitated in a **non-decimated** wavelet representation

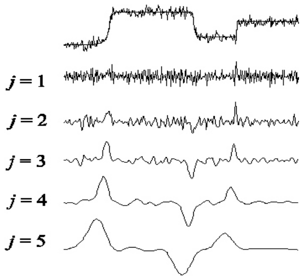


Resolution scale

# Denoising by singularity detection




Rate of increase of the modulus of the wavelet transform across scales is proportional to the local Lipschitz regularity

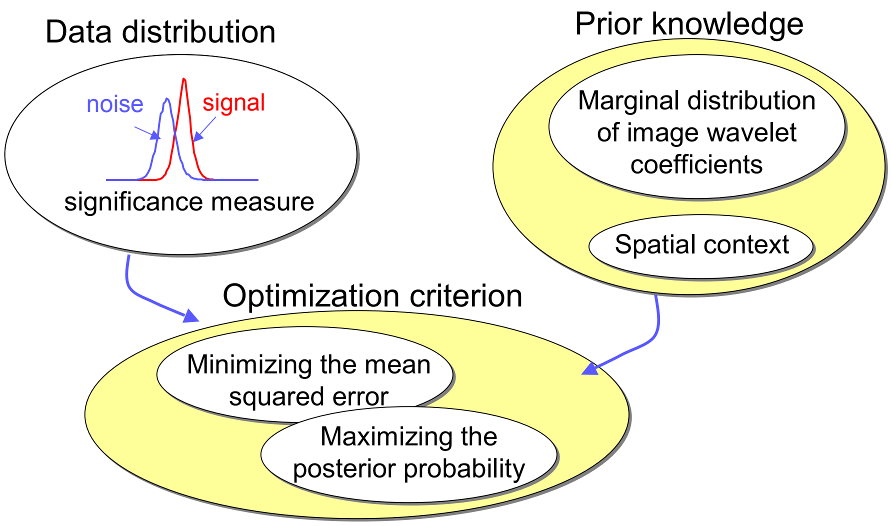




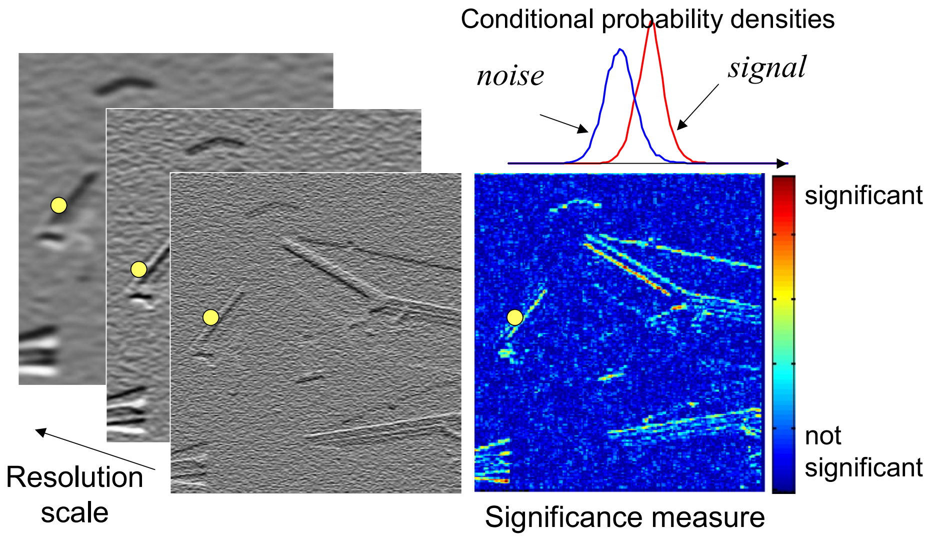




# Bayesian Approach



# Significance Measures

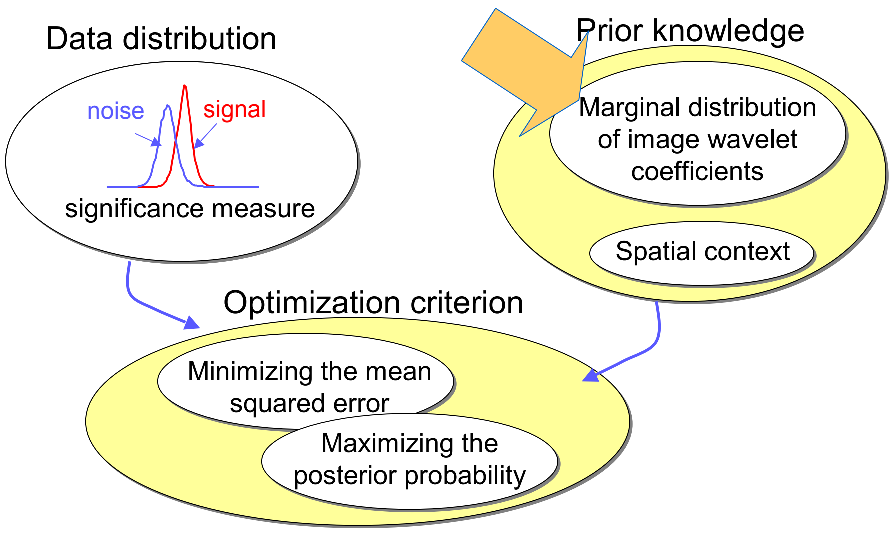






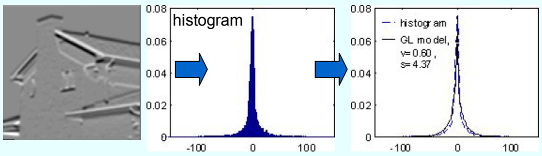


# Bayesian Approach



# Marginal priors

**Generalized Laplacian** (generalized Gaussian) distribution

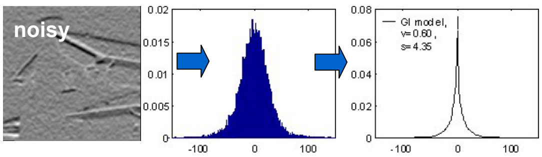



$f(y) = A \exp(-|y/s|^{\nu})$

$s$: scale parameter

$\nu$: shape parameter

$(0 \leq \nu \leq 1)$



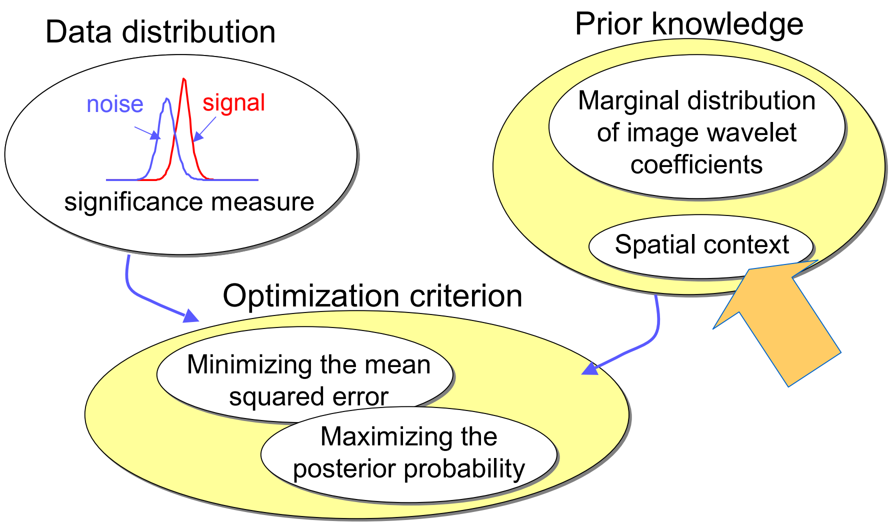

Parameters accurately estimated from a signal corrupted by additive white Gaussian noise

Special case $\nu=1$: **Laplacian** (**double exponential**) $f(y) = A \exp(-|y/s|)$
analytical tractability, usually minor performance degradations

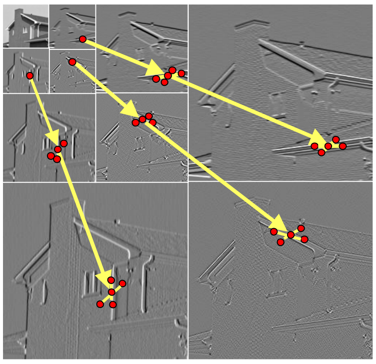





# Bayesian Approach



# Inter- and intrascale dependencies

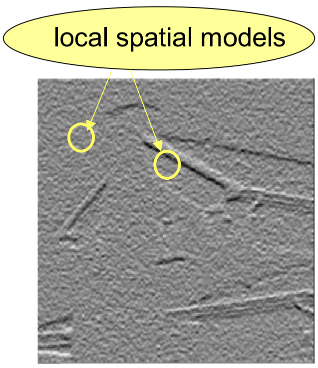


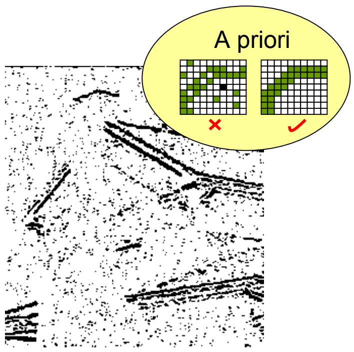
- Bivariate and joint statistics
- Hidden Markov Tree models
- Markov Random Field models
- Priors for context measurements






# Spatial context modeling




- A local variance and locally averaged magnitude are large at the positions of actual edges
- Spatially connected edge clusters are a priori more probable. Markov Random Fields are often used to encode this preference.

# MRF based methods

PSNR=14.9 dB PSNR=24.8 dB PSNR=28.3 dB


Noise free


Input

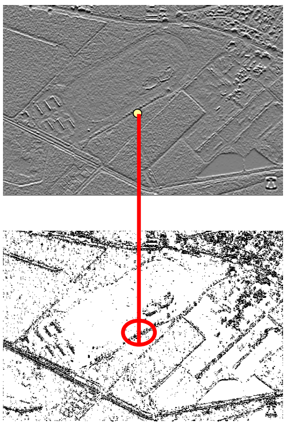

Spatially adaptive Wiener

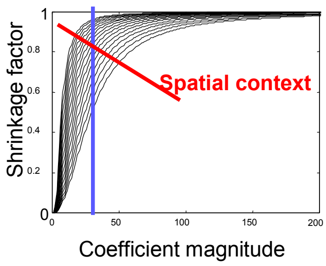

New MRF-based algorithm

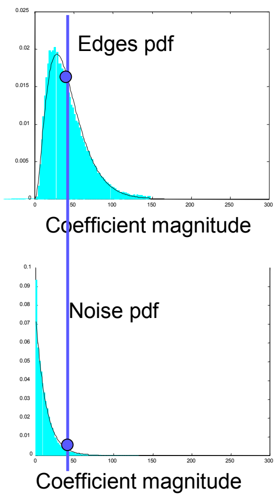




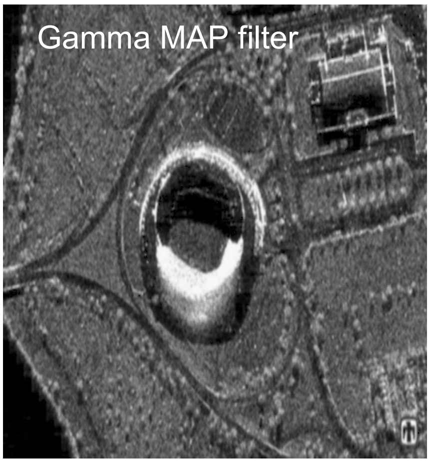
# Adaptive wavelet shrinkage for SAR images

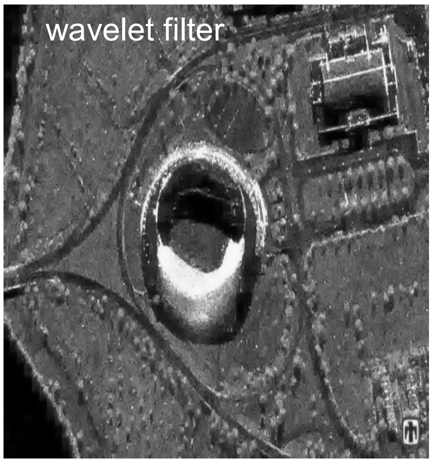




## Despeckling SAR images



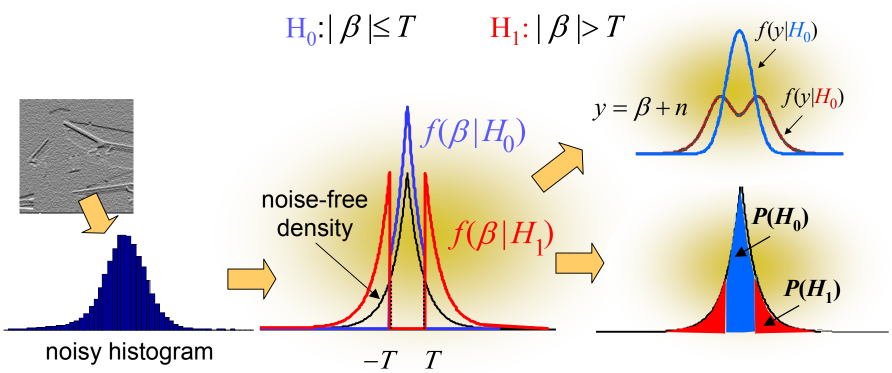






## Statistical modeling: Defining a signal of interest

**Signal of interest**: a noise-free coefficient $\beta$ with magnitude above a given threshold.

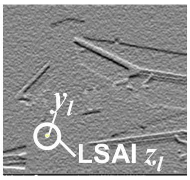


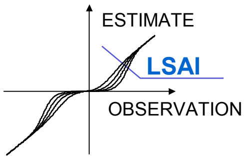
# Locally adaptive denoising: *ProbShrink*

**Pizurica**

## Local spatial activity indicator - LSAI

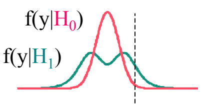
**LSAI – Local Spatial Activity Indicator**



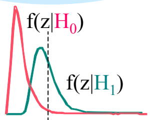




$$y = \beta + n, \qquad \hat{\beta} = P(H_1 \mid y, z)\, y = \frac{\eta \xi \mu}{1 + \eta \xi \mu}\, y$$

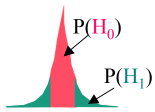
$$\eta = \frac{f(y \mid H_1)}{f(y \mid H_0)} \qquad \xi = \frac{f(z \mid H_1)}{f(z \mid H_0)} \qquad \mu = \frac{P(H_1)}{P(H_0)}$$




noisy coefficient $y$




LSAI $z$




subband statistics






# Multispectral *ProbShrink*

# Multispectral *ProbShrink*





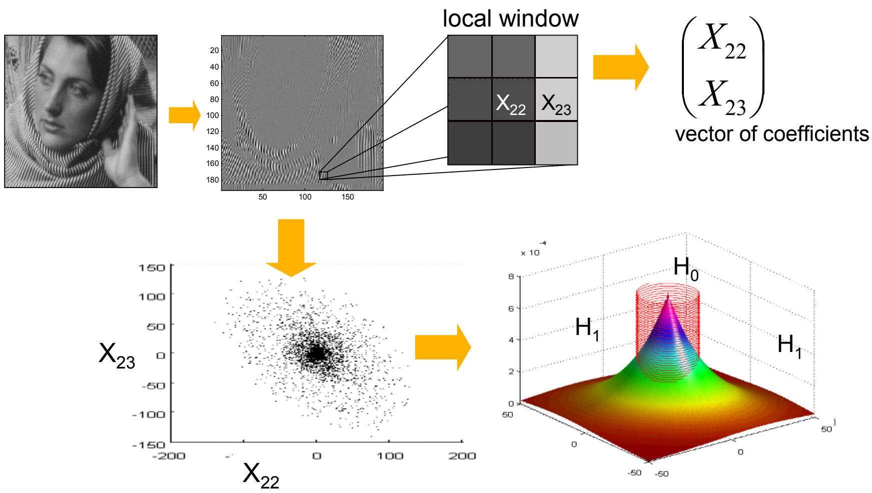



# Different noise level in RGB channels

$\sigma_R = 55, \sigma_G = 25, \sigma_B = 10$


*ProbShrink-SB-YUV*

# *ProbShrink* for correlated noise



# ProbShrink for correlated noise