Digital Image Processing

14 December 2006

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

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

Filters adapted to local noise statistics

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 \( \sigma_f^2 + \sigma_n^2 = \sigma_n^2 \) and if \( n \) is zero mean process: \( \hat{f}(x, y) = g(x, y) \)

What we need to know to proceed?

1. Assuming \( f \) (and \( g \)) is ergodic process: \( \bar{g}(x, y) = m_L \), \( \sigma_g^2 = V_L \)
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 $\sigma$ of AWGN noise:
- **block-based**: use regions (blocks) with the least signal variation
- **smoothing-based**: estimate $\sigma$ 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 $\sigma$

Suppression of multiplicative noise

Two usual approaches:

Input image $I$

$$\log (I+1)$$

Filter for additive noise

$J$

$$\exp(J)-1$$

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 using 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 \( g = f + n \)

Data distribution

Prior knowledge

Optimization criterion

- Minimizing the mean squared error
- Maximizing the posterior probability

\[ p(g \mid f) \]

Bayesian approach using prior knowledge about the unknown image

Denote \( g = f + n \)

Data distribution

Prior knowledge

Optimization criterion

- Minimizing the mean squared error
- Maximizing the posterior probability

\[ p(g \mid f) \]

Maximum a posteriori (MAP) estimate:

\[
\hat{f}_{MAP} = \arg \max_{f \in \mathcal{F}} \frac{p(g \mid f)p(f)}{P(g)}
\]

\[
= \arg \max_{f \in \mathcal{F}} p(g \mid f)p(f)
\]

Minimum mean squared error (MMSE) estimate

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

If \( p(f) \) is Gaussian and noise is also additive Gaussian (then \( p(g \mid f) \) is Gaussian) the MMSE estimate reduces to the Wiener filter.
**Markov Random Field (MRF) prior model**

\[ f = \{f_1, \ldots, f_n\} \]

\[ P(f) = \frac{1}{Z} \exp \left\{ - \sum_{C \in \mathcal{V}} \gamma_C (f) \right\} \]

**clique potentials**

<table>
<thead>
<tr>
<th>neighborhood</th>
<th>cliques</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="cliques" /></td>
</tr>
</tbody>
</table>

Example: penalize isolated peaks

- Negative potential
- Positive potential

**Ising model: an MRF model for binary images**

Pair-site clique potential

\[ V_2(f_i, f_k) = \begin{cases} 
-\gamma, & f_i = f_k \\
\gamma, & f_i \neq f_k 
\end{cases}, \gamma > 0 \]

Samples form the Ising model

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:
  \[ V_2(f_i, f_k) = \alpha (f_i - f_k)^2 \]
  → This model tends to oversmooth the image (destroys edges).

- An example of a discontinuity adaptive model:
  \[ V_2(f_i, f_k) = \begin{cases} 
  \alpha (f_i - f_k)^2, & \text{if } |f_i - f_k| < C \\
  \text{const} (= \alpha C^2), & \text{otherwise}
  \end{cases} \]
  → 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:
\[ \hat{f} = \arg \max_{f \in \mathcal{F}} p(f | g) = \arg \max_{f \in \mathcal{F}} p(g | f) P(f) \]

For white Gaussian noise with zero mean and standard deviation \( \sigma \):
\[ p(g | f) = A \exp\left(- \sum_l (g_l - f_l)^2 / 2\sigma^2\right) \]

MRF prior:
\[ p(f) = Z \exp\left(- \sum_{c \in C} V_c(f)\right) = Z \exp\left(- \sum_{k,l} V_2(f_k, f_l)\right) \]

\[ \hat{f} = \arg \max_{f \in \mathcal{F}} \log(p(g | f) P(f)) = \arg \min_{f \in \mathcal{F}} E(f | g) \]

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

\( \lambda \) is a constant
**Image denoising using MRF priors**

Assume:

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

MAP solution: minimize the posterior energy \( E(f|g) \)

\[
E(f|g) = \sum_{l}(g_l - f_l)^2 + \lambda \sum_{k<l} V_2(f_k, f_l)
\]

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 global minimum (example: Metropolis sampler)
- Genetic algorithms

---

**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 \( E_{TV} \):

\[
E_{TV} = \iint (f(x,y) - g(x,y))^2 + \lambda |\nabla f(x,y)| \, dx \, dy
\]

The assumed underlying image model is in this case a “cartoon model”

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

\[
\begin{align*}
E(f|g) & = \sum_{l}(g_l - f_l)^2 + \lambda \sum_{k<l} V_2(f_k, f_l) \\
\end{align*}
\]
Total variation (TV): a practical method

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

$$\frac{1}{\Omega} \iint (\hat{f}(x,y) - g(x,y))^2 \, dx \, dy = \sigma_n^2$$

$\Omega$ 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 too high or too small 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

⇒ Grass is blurred

Applications:

- Noise reduction in images without texture
- Pre-processing for segmentation
Wavelet domain denoising

Noisy input

Wavelet transform

\[ s_L, \begin{bmatrix} w_1 \\ w_2 \\ w_L \end{bmatrix} \]

scaling coefficients

Remove noise

Estimated noise-free wavelet coefficients

Denoised output

Inverse wavelet transform
Noise model and notation

image domain: \( v_i = f_i + \xi_i; \quad \xi_i \sim N(0, \sigma_{in}) \)

wavelet domain:

\[
\begin{align*}
\sigma^2_j &= S_j^d \sigma_{in}^2 \\
S_{j}^{LH,HL} &= \left( \sum_k g_k^2 \right) \left( \sum_k h_k^2 \right)^{2j-1} \\
S_{j}^{HH} &= \left( \sum_k g_k^2 \right) \left( \sum_k h_k^2 \right)^{2(j-1)}
\end{align*}
\]

\( g_k \) 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
- Median Absolute Deviation (MAD) estimator

\[
\hat{\sigma} = \text{Median}(|w_{1}^{HH}|) / 0.6745
\]
Noise model \( w = y + n; \quad n \sim N(\theta, \sigma) \)

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

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

One can show that

\[
E\left[ (y - w \theta(y))^2 \right] = y^2 (1 - \theta(y))^2 + \sigma^2 \theta(y)^2
\]

The estimator \( \theta_{\text{ideal}}(y) \) that minimizes this error is

The ideal attenuator is

\[
\theta_{\text{ideal}}(y) = \frac{y^2}{y^2 + \sigma^2}
\]

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

\[
\theta_{\text{ideal/soft}}(y) = \begin{cases} 
1, & \text{if } |y| > \sigma \\
0, & \text{if } |y| \leq \sigma 
\end{cases}
\]

oracle thresholding

---

Denoising by wavelet thresholding

Wavelet coefficient values

Thresholded values

Hard thresholding “keep or kill”

Soft thresholding “shrink or kill”
Denoising by wavelet thresholding

\[ w = y + n \]
\[ y \] – noise-free coefficient;
\[ n \] – i.i.d. Gaussian noise

\[ \hat{y}_w = \begin{cases} 0, & |w| < T \\ w, & |w| \geq T \end{cases} \]
Hard thresholding “keep or kill”

\[ \hat{y}_w = \begin{cases} 0, & |w| < T \\ \text{sgn}(w)(|w| - T), & |w| \geq T \end{cases} \]
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

Non-decimated WT

Algorithm à trous

Also called overcomplete and stationary wavelet transform

Inter-scale comparisons are facilitated in a non-decimated wavelet representation
Denoising by singularity detection

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

Prior knowledge

Marginal distribution of image wavelet coefficients

Spatial context

Data distribution

significance measure

noise

signal

Optimization criterion

Minimizing the mean squared error

Maximizing the posterior probability

Bayesian Approach
Significance Measures

Resolution scale

Conditional probability densities

Significance measure

Signal

Noise

Prior knowledge

Marginal distribution of image wavelet coefficients

Spatial context

Optimization criterion

Minimizing the mean squared error

Maximizing the posterior probability

Bayesian Approach

Data distribution

Significance measure

Signal

Noise

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

Data distribution

- Noise
- Signal
- Significance measure

Prior knowledge

- Marginal distribution of image wavelet coefficients
- Spatial context

Optimization criterion

- Minimizing the mean squared error
- Maximizing the posterior probability
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**

Shrinkage factor

Coefficient magnitude

Spatial context

Edges pdf

Noise pdf
Despeckling SAR images

Statistical modeling: Defining a signal of interest

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

$H_0$: $|\beta| \leq T$  
$H_1$: $|\beta| > T$

Noisy histogram

- $f(\beta|H_0)$
- $f(\beta|H_1)$

$P(H_0)$
$P(H_1)$
Locally adaptive denoising: **ProbShrink**

**Local spatial activity indicator - LSAI**

\[ y = \beta + n, \quad \hat{\beta} = P(H_1 | y, z)^{y} = \frac{\eta \xi \mu}{1 + \eta \xi \mu} y \]

\[ \eta = \frac{f(y | H_1)}{f(y | H_0)} \]

\[ \xi = \frac{f(z | H_1)}{f(z | H_0)} \]

**Multispectral ProbShrink**

![Multispectral ProbShrink](image)
**Multispectral ProbShrink**

Different noise level in RGB channels

\[ \sigma_R = 55, \sigma_G = 25, \sigma_B = 10 \]
**ProbShrink for correlated noise**

- Local window
- Vector of coefficients
- Probability density functions $H_0$ and $H_1$
ProbShrink for correlated noise

input  ProbShrink for white noise  ProbShrink for colored noise