Lecture 15:

Nonlinear Filters II:
Wavelet Shrinkage, Bilateral Filtering, NL-Means

Contents

M I A
1 2
3 4
5 6
7 8
9 10
11 12
13 14
15 16
17 18
19 20
21 22
23

Motivation

- Linear shift invariant lowpass filters can denoise images (Lecture 11).
  However, they destroy important information by blurring edges.
- Median filters offer edge-preserving denoising (Lecture 14).
  They work well for impulse noise.
  However, they are less powerful for other noise types such as Gaussian noise.
  Moreover, simple implementations are slow for large masks.
- Are there other nonlinear filters that are better for Gaussian noise or faster?

Wavelet Shrinkage

Procedure in Three Steps (Donoho / Johnstone 1992)

- *Analysis Step:*

  Represent the image in a wavelet basis (Lecture 7):
  - The useful signal part is well represented by a few wavelet coefficients with large magnitude (sparsity).
  - Typical noise such as moderate Gaussian noise is present in numerous wavelet coefficients with small magnitude.

- *Shrinkage Step:*

  Eliminate noise by shrinking wavelet coefficients with small magnitude towards 0. (This is the only nonlinear step. Do not shrink the scaling coefficient !!)

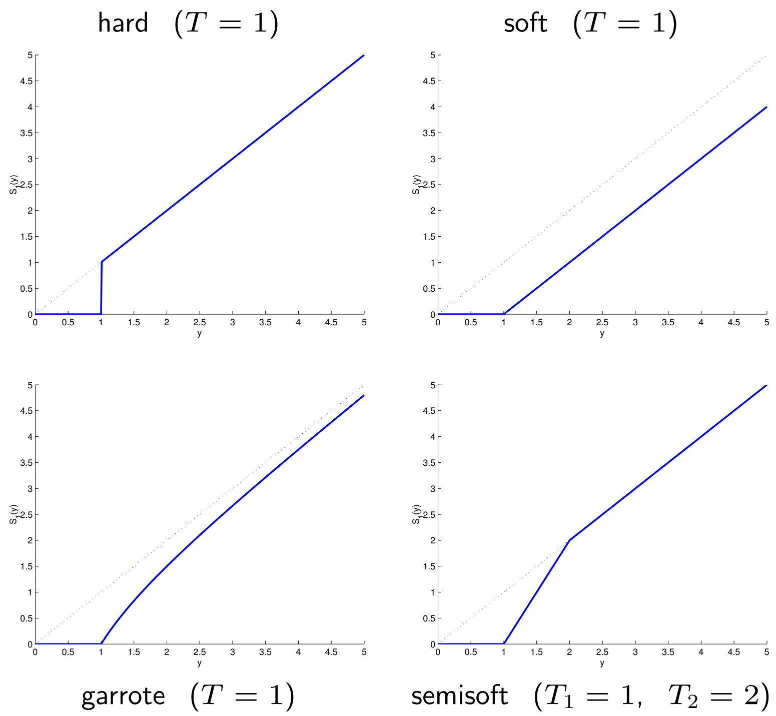
- *Synthesis Step:*

  Reconstruct the image from the modified coefficients.

This procedure also resembles image compression with wavelets (Lecture 7).
It is very fast due to the linear complexity of the Fast Wavelet Transform.

M I A
1 2
3 4
5 6
7 8
9 10
11 12
13 14
15 16
17 18
19 20
21 22
23

Graphical Representation of Different Shrinkage Strategies



Different shrinkage functions. Horizonal axis: original coefficient. Vertical axis: shrunken coefficient. Author: P. Mrázek.

Wavelet Shrinkage (3)

Four Popular Strategies for Shrinking a Wavelet Coefficient d towards 0:

Hard Shrinkage:

- can clearly tell good from bad
- assumes that large coefficient magnitudes respresent correct signal, while small magnitudes are caused by noise:

$$S_T(d) \;:=\; \begin{cases} 0 & (|d| \le T), \\ d & (|d| > T). \end{cases}$$

Soft Shrinkage:

- Everybody is bad by the same degree.
- assumes that every coefficient contains the same amount of noise:

$$S_T(d) \;:=\; \begin{cases} 0 & (|d| \le T), \\ d - \operatorname{sgn}(d)\, T & (|d| > T). \end{cases}$$

Wavelet Shrinkage (4)

Garrote Shrinkage:

- There is a gradual transition between good and bad.
- The larger the magnitude is, the more likely it is caused by the signal:

$$S_T(d) \;:=\; \begin{cases} 0 & (|d| \le T), \\ d - \frac{T^2}{d} & (|d| > T). \end{cases}$$

Semisoft Shrinkage (Firm Shrinkage):

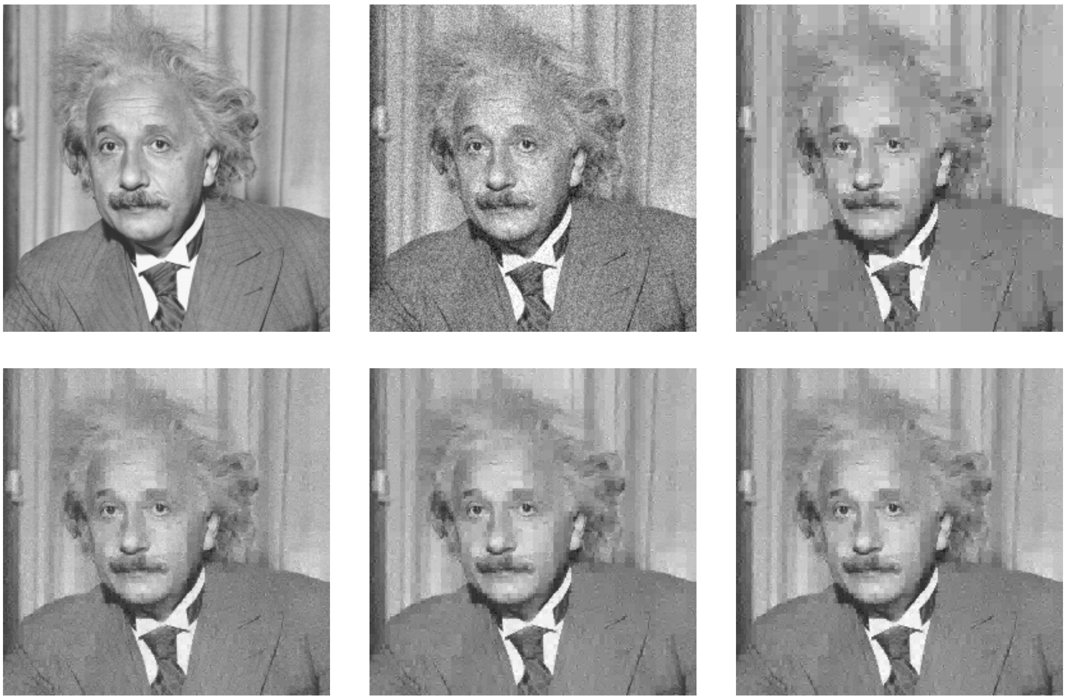
- Some are always good, some are always bad, and there's a greyzone inbetween.
- two thresholds T_1, T_2 with linear transition zone:

$$S_{T_1,T_2}(d) \;=\; \begin{cases} 0 & (|d| \le T_1), \\ d - \operatorname{sgn}(d)\, T_1 \frac{T_2 - |d|}{T_2 - T_1} & (T_1 < |d| \le T_2), \\ d & (|d| > T_2). \end{cases}$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Wavelet Shrinkage (5)

original, 256×256 | Gaussian noise, $\sigma = 17.3$ | hard, $T = 50$



soft, $T = 18$ | garrote, $T = 29$ | semisoft, $T_1 = 24$, $T_2 = 135$

Denoising performance of the different strategies for Haar wavelet shrinkage. The shrinkage parameters are chosen such that the PSNR is maximised. Author: J. Weickert.

M I A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Wavelet Shrinkage (6)

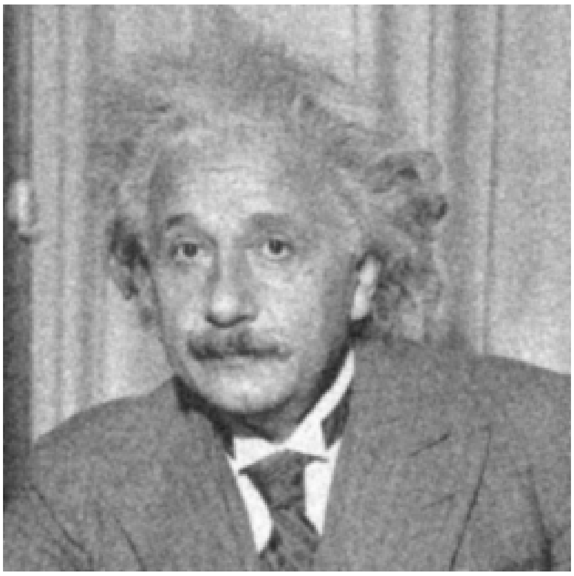
Important Remarks

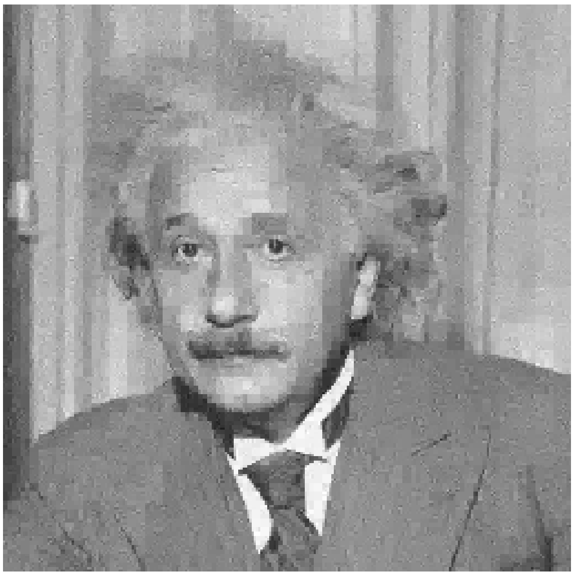
- Wavelet shrinking only affects the wavelet coefficients.
  The scaling coefficient carries information on the average grey value.
  Gaussian noise with zero mean is not supposed to alter it.
  Thus, do not shrink the scaling coefficient !!
- Classical wavelet shrinkage does not prefer certain frequencies.
  This is in contrast to linear Fourier-based filtering.
  The shrinkage operation treats all wavelet coefficients equally:
  – It only depends on the *magnitude* of the wavelet coefficient.
  – However, it does not depend on their *scale (frequency)*.

Wavelet Shrinkage (7)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Why Does Wavelet Shrinkage Handle Edges Better than Linear Filtering ?





Comparison between a linear filter and wavelet shrinkage. **Left:** Optimal denoising by Gaussian convolution ($\sigma = 0.75$). Edges are blurred. **Right:** Optimal denoising with semisoft Haar wavelet shrinkage ($T_1 = 24$ und $T_2 = 135$). Edges are preserved more faithfully. Author: J. Weickert.

Wavelet Shrinkage (8)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Linear Shift Invariant Filtering

- The Fourier basis is not localised at all.
  Thus, each Fourier coefficient has a *global* impact on the entire image.
- Linear shift invariant denoising eliminates or damps certain Fourier coefficients.
- Thus, an LSI filter cannot be tuned to treat important local structures such as edges differently than noise somewhere else.

Wavelet Shrinkage

- The wavelet basis benefits from its localisation.
  The impact of a single wavelet coefficient is *local.*
- Edges create wavelet coefficients with large magnitude.
- Hard or semisoft shrinkage does not change them at all.
  Garrote shrinkage damps them only a little bit.
  Soft shrinkage affects only a small proportion of their total magnitude.
- Thus, edges are well preserved.

Pseudo-Gibbs Artifacts

- Wavelet shrinkage can behave in an undesired way:
  It can increase the largest grey value and decrease the smallest one!
- In particular near edges, such over- and undershoots appear frequently.
  They are called *pseudo-Gibbs artifacts.*
- Example: Representing the signal $(10,1,2,1)^\top$ in its Haar wavelet basis gives

$$\begin{pmatrix} 10 \\ 1 \\ 2 \\ 1 \end{pmatrix} = 7 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} + 4 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} + \frac{9}{\sqrt{2}} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

After hard wavelet shrinkage with $T = 5$ we obtain the filtered signal

$$7 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} + 0 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} + \frac{9}{\sqrt{2}} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \\ 3.5 \\ 3.5 \end{pmatrix}.$$

Further Properties of Wavelet Shrinkage

- It does not affect the average grey value:
  - Wavelets have mean 0. Shrinking them does not change the mean.
  - The scaling coefficient – which represents the mean – is unaltered.
- Sometimes people use different shrinkage parameters for different scales.
  This introduces some frequency dependence.
  However, it is more difficult to handle these multiple parameters.
- Usual wavelet shrinkage is not shift invariant and not invariant under rotations.

How to Make Wavelet Shrinkage Shift-Invariant

- simple averaging over all possible translations at all scales
  (cycle spinning, algorithm a trous)
- equivalent to a wavelet-like representation without up- and downsampling steps
  (undecimated wavelet transform, stationary wavelet transform)
- reduces also pseudo-Gibbs artifacts
- further improvements possible by iterating
- quality gain at the expense of a lower efficiency

M I A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Bilateral Filtering

(Aurich/Weule 1995, Smith/Brady 1997, Tomasi/Manduchi 1998)

- discrete adaptive averaging with weights that
  - decrease with increasing distance from the stencil centre $\boldsymbol{x}_i$:
    spatial weights $w\,(|\boldsymbol{x}_i - \boldsymbol{x}_j|)$
  - decrease with increasing distance of the grey values (preserves edges):
    tonal weights $g\,(|f_i - f_j|)$
- computes filtered signal / image $\boldsymbol{u}$ of $\boldsymbol{f}$ via

$$u_i \;=\; \frac{\sum_j \, g\,(|f_i - f_j|)\;\, w\,(|\boldsymbol{x}_i - \boldsymbol{x}_j|)\;\, f_j}{\sum_j \, g\,(|f_i - f_j|)\;\, w\,(|\boldsymbol{x}_i - \boldsymbol{x}_j|)}$$

  with decreasing nonnegative functions w and g
- typical functions for w and g: Gaussians with standard deviation σ_s and σ_t
- normalisation by denominator guarantees that all weights sum up to 1

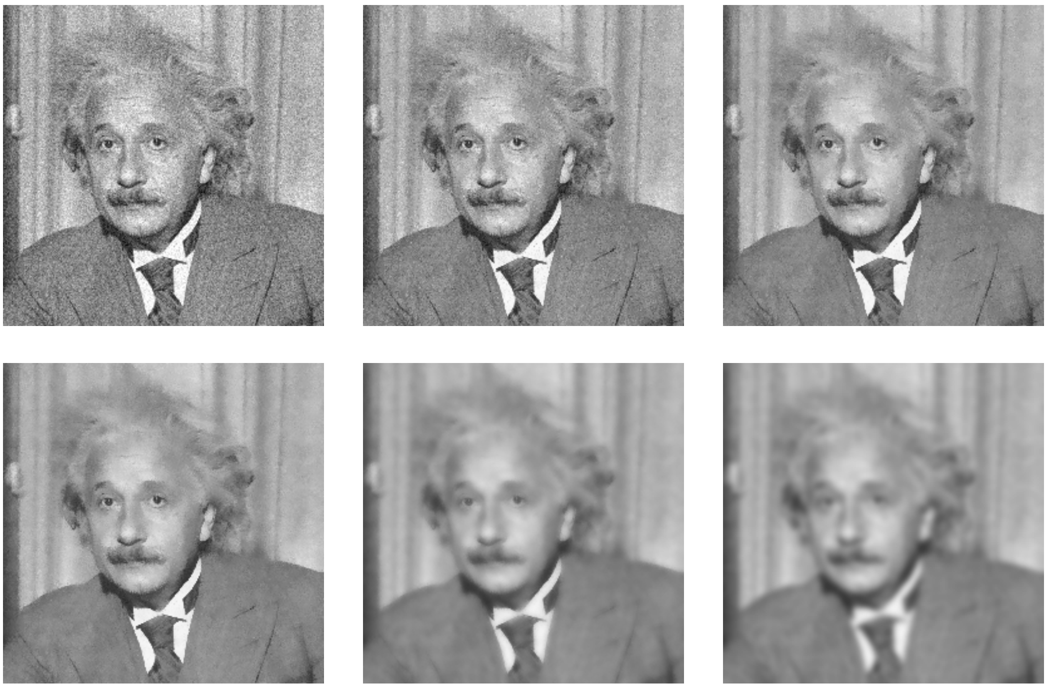
Bilateral Filtering (2)

Left: Original image, 325×356 pixels. **Right:** After bilateral filtering. Authors: Tomasi / Manduchi.

Bilateral Filtering (3)

MIA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

original, 256×256 — Gaussian noise, $\sigma = 17.3$ — filtered, $\sigma_t = 15$

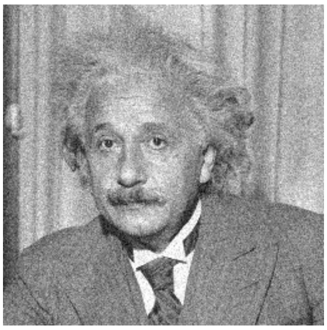


filtered, $\sigma_t = 25$ — filtered, $\sigma_t = 100$ — filtered, $\sigma_t = 1000$

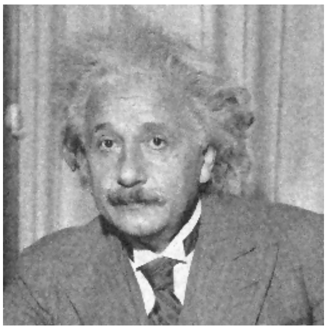
Influence of the tonal parameter σ_t on the bilateral filtering result ($\sigma_s = 2.5$). Note that for large σ_t, the tonal weight tends to 1 and one approximates Gaussian smoothing. Author: J. Weickert.

Bilateral Filtering (4)

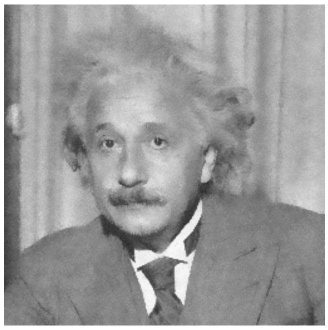
original, 256×256

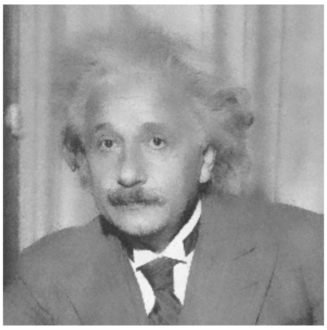


Gaussian noise, $\sigma = 17.3$

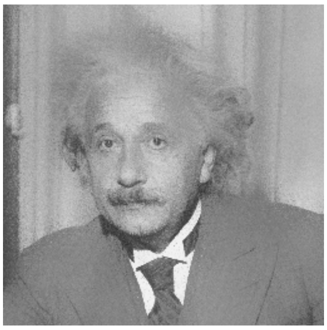


filtered, $\sigma_s = 1$

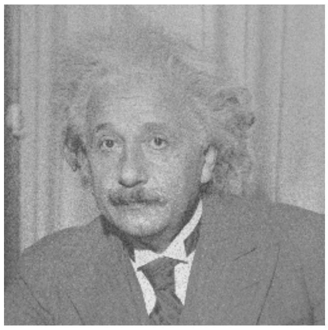




filtered, $\sigma_s = 4$



filtered, $\sigma_s = 10$



filtered, $\sigma_s = 50$

Influence of the spatial parameter σ_s on the bilateral filtering result ($\sigma_t = 35$). A large σ_s leads to pure tonal averaging: Structures remain sharp, while the contrast decreases globally. Author: J. Weickert.

Bilateral Filtering (5)

Properties of Bilateral Filtering

- shift-invariant and invariant under rotations
- no over- and undershoots: satisfies *maximum–minimum principle*

$$\min_i f_i \;\leq\; u_j \;\leq\; \max_i f_i \qquad \forall\, j.$$

- can be iterated
- slower than wavelet shrinkage (in particular for large σ_s)
- does not preserve the average grey value
- fairly popular in computer graphics
- requires to specify two parameters σ_s and σ_t
- can give results with good quality, if both parameters are chosen well

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

NL Means

(Buades et al. 2005)

- adaptive averaging of pixels that belong to similar patches within the entire image
- Let $\mathcal{N}_i$ be a square-shaped $(2m+1) \times (2m+1)$ neighbourhood around a pixel i.
- The grey value distance between two neighbourhoods $\mathcal{N}_i$ and $\mathcal{N}_j$ is expressed by the Euclidean distance between the corresponding vectors in $\mathbb{R}^{(2m+1)^2}$:

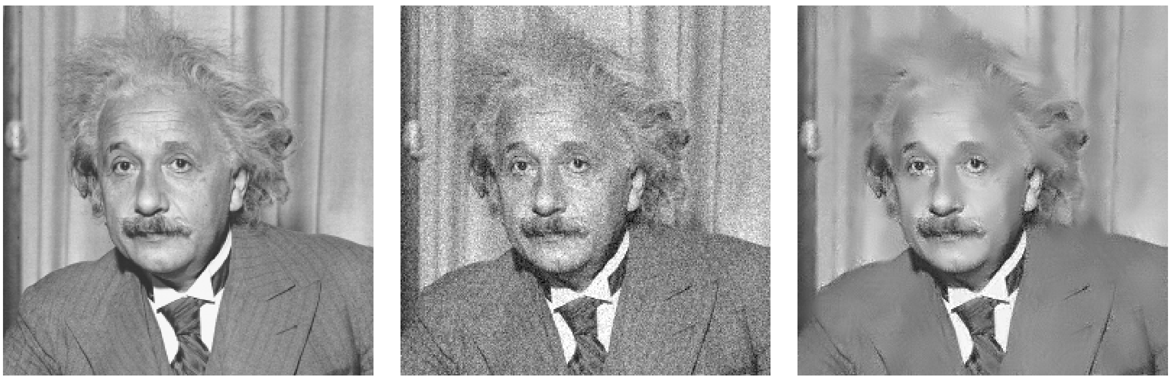
$$|\boldsymbol{f}(\mathcal{N}_i) - \boldsymbol{f}(\mathcal{N}_j)| \, .$$

- The *nonlocal (NL) means* algorithm performs weighted averaging where the more similar patches have higher weights:

$$u_i \;=\; \frac{\sum_j g\left(|\boldsymbol{f}(\mathcal{N}_i) - \boldsymbol{f}(\mathcal{N}_j)|\right) f_j}{\sum_j g\left(|\boldsymbol{f}(\mathcal{N}_i) - \boldsymbol{f}(\mathcal{N}_j)|\right)} \, .$$

- Often a Gaussian is used as weight function g.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23



Left: Original image, 256×256 pixels. **Middle:** After adding Gaussian noise with $\sigma = 17.3$. **Right:** After filtering with NL-means with optimised parameters (7×7 neighbourhood, search window restricted to 15×15 pixels, Gaussian weight function with $\sigma = 80$). Author: J. Weickert.

NL Means (3)

Properties of NL Means Filtering

- satisfies a maximum-minimum principle
- shift invariant
- not rotationally invariant for square-shaped patches;
  easy to cure with disk-shaped patches
- does not preserve the average grey value
- slow due to a excessive comparison of patches
- Usually one restricts the search space to a window around pixel i.
  This speeds up the method and hardly influences the quality of the results.
- three parameters: search window size, patch size, Gaussian standard deviation
- can give results of very high quality for suitable parameters

Summary

Summary

- Wavelet shrinkage eliminates wavelet coefficients with small magnitude.
  The scaling coefficient (representing the average grey value) is untouched.
- four most important shrinkage functions: hard, soft, garrote, semisoft.
- The locality of wavelets allows the preservation of edges.
- Bilateral filtering performs adaptive averaging with spatial and tonal weights.
- NL means comes down to nonlocal adaptive averaging of pixels that belong to similar patches.

References

- W. Bäni: *Wavelets*. Oldenbourg, München, 2002.
  (fairly simple introduction to wavelet concepts; in German)
- S. Mallat: *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, Second Edition, 1999.
  (wavelet bible)
- I. K. Folder, C. Kamath: Denoising through wavelet shrinkage: an empirical study. *Journal of Electronic Imaging*, Vol. 12, No. 1, 151–160, Jan 2003.
  (`https://pdfs.semanticscholar.org/37cf/f1c99b702b716896c12f86b589e83d20fc0e.pdf`)
  (evaluates different shrinkage strategies)
- V. Aurich, J. Weule: Non-linear Gaussian filters performing edge preserving diffusion. In G. Sagerer, S. Posch, F. Kummert (Eds): *Mustererkennung 1995*. Springer, Berlin, pp. 538–545, 1995.
  (early paper on bilateral filtering)
- S. M. Smith, J. M. Brady: SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, Vol. 23, No. 1, 45-78, 1997.
  (reinvention of the bilateral filter)
- C. Tomasi, R. Manduchi: Bilateral filtering for gray and color images. *Proc. Sixth International Conference on Computer Vision* (ICCV 1998, Bombay, India, Jan. 1998), pp. 839–846.
  (`https://users.soe.ucsc.edu/~manduchi/Papers/ICCV98.pdf`)
  (made bilateral filtering popular)
- A. Buades, B. Coll, J.-M. Morel: A non-local algorithm for image denoising. *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2005, San Diego, CA, June 2005), pp. 60–65.
  (introduced NL means)