Bivariate shrinkage functions for wavelet based denoising exploiting interscale dependency

**Denoising**

- Image or video denoising is the process of removing noise from a signal
- It has many applications
  - Medical signal/image analysis (ECG, CT, MRI etc.)
  - Data mining – process of extracting patterns from data; data mining is about analyzing data, about extracting information out of data
- Example of denoised image

![Example of denoised image](image)

- The left is a noisy MR brain image. Before the analysis, we would like to clean it up
- Useful for better interpretation of the image
- Denoising implies multiscale decomposition: breaking into basic components at multiple scales and using properties from one level to another
- Multiscale decomposition is used extensively in image processing
- The most common problem in signal processing is a natural image corrupted by Gaussian noise
- Steps in removing noise
  - 1) Calculate the wavelet transform of the noisy signal.
  - 2) Modify the noisy wavelet coefficients according to some rule.
  - 3) Compute the inverse transform using the modified coefficients.
- **Thresholding**
  - Denoising by thresholding in the wavelet domain has been developed principally by Donoho who introduced RiskShrink with the minimax threshold, VisuShrink with the universal threshold, and discussed both hard and soft thresholds
  - RiskShrink and the minimax threshold. The threshold is calculated in order to minimize the possible loss while maximizing the potential gain
  - VisuShrink – the universal threshold represents the degree of improbability below which a specific event of that probability cannot be attributed to chance
  - Soft threshold = smooth transitions between the original and the detected values
  - Hard threshold = values above the threshold are kept, values below are deleted
  - **David Leigh Donoho**, born on March 5, 1957 in Los Angeles, is a professor of mathematics and statistics at Stanford University
  - His work includes multiscale geometric analysis, developments of wavelets for denoising and compressed sensing
  - The main idea is to report coefficients to the threshold value $T$: subtract value $T$ from all coefficients larger than $T$ and to set all the rest to zero
- Statistical approaches: The basic idea is to model wavelet transform coefficients with prior probability distributions.
- The solution to this problem is the MAP estimator - requires a priori knowledge about the distribution of wavelet coefficients.
- The MAP estimator is used in Bayesian statistics

**BAYESIAN DENOISING**

- In this section, the denoising of an image corrupted by white Gaussian noise will be considered:
  - \( g = x + n \)
    - where \( g \) = signal with noise
    - \( x \) = desired signal
    - \( n \) = independent Gaussian noise
- The quest is to estimate \( x \), the desired signal, the useful information
- By using orthogonal wavelet transform, the problem can be formulated as
  - \( y = w + n \)
    - \( y \) = noisy wavelet transformation
    - \( w \) = true coefficient
    - \( n \) = noise
- The goal is to determine \( w \), by using MAP estimator

**Marginal Model**

- The classical MAP estimator
  \[ \hat{w}(y) = \arg \max_w p_{yw}(w) \]
- By using the probability distribution functions pdf of the noise \( p_n \) and of the signal coefficient \( p_w \), the formula can be rewritten:
  \[ \hat{w}(y) = \arg \max_w \left[ p_{yw}(y | w) \cdot p_w(w) \right] \]
  \[ = \arg \max_w \left[ p_n(y - w) \cdot p_w(w) \right] \]

- Problems
  - It can be difficult to estimate the parameters for a specific image, especially from noisy data the estimators for these models may not have simple closed form solution and can be difficult to obtain. The solution for these problems usually requires numerical techniques

- Examples

  - Illustrates the histogram of the wavelet coefficients computed from several natural images.
- By further developing MAP estimator and rewriting formulas for the MAP estimator an equivalent for the previous formula is obtained:

$$\hat{w}(y) = \arg \max_w \left[ \log(p_w(y - w)) + \log(p_y(w)) \right]$$

- If $p_w$ is Laplacian then we can obtain the soft shrinkage function

- The Laplacian pdf and corresponding shrinkage function

- The distribution functions

- Shrinkage function

- **Bivariate Models**
  - Marginal models cannot model statistical dependencies between wavelet coefficients
  - There are strong dependencies between neighbor coefficients such as between a coefficient, its parent and their siblings
  - By modifying the Bayesian estimator to take into account the statistical dependency between a coefficient and its parent, the equation can be written as follows:

$$\hat{w}(y) = \arg \max_w \left[ p_{yw}(y \mid w) \cdot p_w(w) \right]$$

$$\hat{w}(y) = \arg \max_w \left[ p_n(y - w) \cdot p_w(w) \right]$$

- In order to use this equation to estimate the original signal both probabilities of distribution must be known and assume the noise independent and identically distributed gaussian
- Problems
- The same problem as in marginal case appears. What kind of joint pdf models the wavelet coefficients? The joint empirical coefficient-parent histogram can be used to observe $p_w(w)$
- Examples
- For this purpose here is an example
- Empirical joint parent-child histogram of wavelet coefficients (computed from the Corel image database)

Joint shrinkage function derived numerically from the empirical joint parent-child histogram
- **APPLICATION TO IMAGE DENOISING**

- Lena image is used

- a critically sampled orthogonal discrete wavelet transform, with Daubechies length-8 filter

- Zero mean white Gaussian noise is added to the original image

- 1. Original image
   2. Noisy image with PSNR = 20.02 dB
   3. Denoised image using soft thresholding; PSNR = 27.73 dB
   4. Denoised image using new bivariate shrinkage function given in (30); PSNR = 28.83 dB.

- **Conclusions**

- new bivariate (having 2 variables) distributions are proposed for wavelet coefficients of natural images in order to characterize the dependencies between a coefficient and its parent

- the corresponding bivariate shrinkage functions are derived from them using Bayesian estimation, in particular, the MAP estimator.