

Discrete Wavelet Diffusion for Image Denoising

Kashif Rajpoot¹, Nasir Rajpoot², J. Alison Noble¹

¹Department of Engineering Science, University of Oxford, UK.

²Department of Computer Science, University of Warwick, UK.

{kashif, noble}@robots.ox.ac.uk, nasir@dcs.warwick.ac.uk

Abstract: Nonlinear diffusion, proposed by Perona and Malik, is a well-known method for image denoising with edge preserving characteristics. Recently, nonlinear diffusion has been shown to be equivalent to iterative wavelet shrinkage, but only for (1) Mallat-Zhong dyadic wavelet transform and (2) Haar wavelet transform. In this paper, we generalize the equivalence of nonlinear diffusion to non-linear shrinkage in the standard discrete wavelet transform (DWT) domain. Two of the major advantages of the standard DWT are its simplicity (as compared to 1) and its potential to benefit from a greater range of orthogonal and biorthogonal filters (as compared to both 1 and 2). We also extend the wavelet diffusion implementation to multiple scales. The qualitative and quantitative results shown for a variety of images contaminated with noise demonstrate the promise of the proposed standard wavelet diffusion.

1 Introduction

A common problem with image or video signals is the contamination with undesired noise. This poses problems both for visual quality and automated analysis. Many automated analysis operations (e.g., segmentation, edge detection) are highly sensitive to noise. It is often desirable to remove or suppress the noise by means of a pre-processing operation. Many denoising methods consider noise as either high-frequency content or outlier, potentially confusing the edges with noise and smoothing them during the process of noise removal. Edges are important features containing structural information that should be preserved. Nonlinear diffusion, initially proposed by Perona-Malik [1], is a popular iterative method for noise removal that attempts to perform edge-preserving smoothing.

During the last decade or so, the discrete wavelet transform (DWT) has also been shown [2] to be a highly successful way of signal denoising, as it effectively separates signal and noise in the wavelet domain. The DWT based denoising, termed also as wavelet shrinkage, works by thresholding (often, nonlinearly) the wavelet coefficients before reconstructing the denoised signal from wavelet decomposition coefficients. One of the problems with wavelet shrinkage is that it results in valleys around edges (termed as *ringing artifacts*) which can be reduced by cycle spinning [3], at the cost of increased computational and storage complexity.

Recently, it was shown in [4, 5] that a single step of nonlinear diffusion can be considered equivalent to a single shrinkage iteration of coefficients of Mallat-Zhong dyadic wavelet transform (MZ-DWT) [6]. In other words, it is possible to perform diffusion in the wavelet domain such that the benefits of both the techniques could be combined. Nonlinear diffusion begins with an edge estimate, typically obtained through a gradient operator, which may be badly influenced by the noise present in the image. It is believed that the edge estimate in the wavelet domain is much less sensitive to noise due to its characteristics of better low- and high-frequency separation. However, MZ-DWT is an uncommon variant of standard DWT and is not

commonly utilized. It has its own subband filtering framework and a set of wavelet filters, which are usually derived from the derivative of a smoothing function (e.g., cubic spline). Moreover, MZ-DWT may not be able to benefit from the great variety of orthogonal and biorthogonal wavelet filters. In another work [7], the diffusivity function of nonlinear diffusion is shown to have equivalence with Haar wavelet shrinkage. Again, the work is restricted only to the use of Haar wavelet filter.

In this paper, we combine the power of nonlinear diffusion and wavelet shrinkage and generalize their equivalence by establishing a relationship between the standard DWT and nonlinear diffusion. The generalization enables one to perform diffusion filtering utilizing the wavelet coefficients obtained from standard DWT, which is simple and commonly used in image processing research. Thus, one can take advantages of both popular denoising methods – wavelets and nonlinear diffusion – in a single wavelet diffusion framework. We hope that this work will make the wavelet diffusion more widely usable while offering the use of a range of wavelet filters.

This paper is organized as follows: next section establishes the relationship between DWT and nonlinear diffusion after briefly introducing them. Section 3 provides experimental details and presents some qualitative and quantitative results. The paper ends with a brief discussion of results, conclusions and future directions.

2 Discrete Wavelet Diffusion

We present a brief overview of Perona-Malik nonlinear diffusion and DWT and later show the equivalence between the two frequently used denoising techniques.

2.1 Nonlinear Diffusion

The nonlinear diffusion filtering operation is governed by the following nonlinear partial differential equation (PDE):

$$\frac{\partial}{\partial t} [f_t(x, y)] = \text{div}[c(x, y) \cdot \nabla f_t(x, y)] \quad (1)$$

where $f_t(x, y)$ is the initial noisy image at time t and the solution of the above PDE iteratively yields a filtered version which can be considered as the denoised image. In the above, $c(x, y)$ is the diffusion coefficient controlling the diffusion amount and is typically described by a diffusivity function g of the form:

$$c(x, y) = g(|\eta(x, y)|) = \frac{1}{1 + \left(\frac{|\eta(x, y)|}{\lambda}\right)^2} \quad (2)$$

where $|\eta(x, y)|$ is the edge estimate at pixel (x, y) , usually approximated by a gradient-magnitude operator, and λ is the edge threshold parameter. The diffusivity function $g(|\eta|)$ is a nonnegative monotonically decreasing function and, depending upon the value of the edge threshold λ , encourages homogenous regions with reduced noise while preserving the edges. The nonlinear diffusion PDE of (1) can be expanded to:

$$\frac{\partial}{\partial t} [f_t(x, y)] = \frac{\partial}{\partial x} \left[c(x, y) \cdot \frac{\partial}{\partial x} f_t(x, y) \right] + \frac{\partial}{\partial y} \left[c(x, y) \cdot \frac{\partial}{\partial y} f_t(x, y) \right] \quad (3)$$

substituting the time-derivative of $f_t(x, y)$ by its forward difference and discretizing:

$$\frac{f_{t+\Delta t}(x, y) - f_t(x, y)}{\Delta t} = \frac{d}{dx} \left[c(x, y) \cdot \frac{d}{dx} f_t(x, y) \right] + \frac{d}{dy} \left[c(x, y) \cdot \frac{d}{dy} f_t(x, y) \right]$$

Let $\Delta t = 1$ and replacing $c(x, y)$, as defined in (2) above, with $1 - p(x, y)$, we obtain:

$$\begin{aligned} f_{t+1}(x, y) = f_t(x, y) &+ \frac{d^2}{dx^2} f_t(x, y) + \frac{d^2}{dy^2} f_t(x, y) \\ &- \frac{d}{dx} \left[p(x, y) \cdot \frac{d}{dx} f_t(x, y) \right] - \frac{d}{dy} \left[p(x, y) \cdot \frac{d}{dy} f_t(x, y) \right] \end{aligned} \quad (4)$$

Usually, $f_t(x, y)$ is initialized by $f_0(x, y) = f(x, y)$, where $f(x, y)$ is the original noisy image. The solution to the above equation yields a denoised image after a certain number of iterations.

2.2 Discrete Wavelet Transform (DWT)

Wavelets offer a simple and elegant framework for simultaneous time-frequency and multiresolution analysis. In wavelet analysis, a function is decomposed into a superposition of dilated and scaled versions of a mother wavelet basis function. The decomposition highlights interesting features in the wavelet domain which are often not visible in the spatial domain. In DWT, an image is filtered into four subbands at each resolution and the low frequency subband is further filtered through an iterative process to provide the multiresolution representation. Here, we give an overview of the fast implementation of the discrete wavelet transform (DWT) [8]. The *forward wavelet transform* (FDWT) operation (termed also as *decomposition* or *analysis*) is given by:

$$\begin{aligned} A_j(x, y) &= f(x, y) \otimes H(y) \otimes H(x) \\ W_j^H(x, y) &= f(x, y) \otimes H(y) \otimes G(x) \\ W_j^V(x, y) &= f(x, y) \otimes G(y) \otimes H(x) \\ W_j^D(x, y) &= f(x, y) \otimes G(y) \otimes G(x) \end{aligned} \quad (5)$$

where $f(x, y)$ is the given image, \otimes denotes the convolution operation, and H and G are one-dimensional low-pass and high-pass decomposition filters, respectively. The image is decomposed into four subbands; A_j denotes the low-frequency approximation subband and W_j^i , $i \in \{H, V, D\}$ denotes the high-frequency subband at scale j and i being one of the horizontal (H), vertical (V), or diagonal (D) orientations. The original image can be reconstructed through the *inverse wavelet transform* (IDWT) operation (termed also as *reconstruction* or *synthesis*):

$$\begin{aligned} f(x, y) = A_j(x, y) \otimes \tilde{H}(x) \otimes \tilde{H}(y) &+ W_j^H(x, y) \otimes \tilde{G}(x) \otimes \tilde{H}(y) \\ &+ W_j^V(x, y) \otimes \tilde{H}(x) \otimes \tilde{G}(y) + W_j^D(x, y) \otimes \tilde{G}(x) \otimes \tilde{G}(y) \end{aligned} \quad (6)$$

where \tilde{H} and \tilde{G} are the one-dimensional low-pass and high-pass reconstruction filters and derived as conjugate or dual of the decomposition filters H and G depending upon whether the filters are orthogonal or biorthogonal, respectively.

Assuming that the original image $f(x, y)$ is part of the scale-space such that the wavelet coefficients can be manipulated after decomposition, the original image and the reconstructed image can be considered as $f_t(x, y)$ and $f_{t+1}(x, y)$, respectively. Using this convention and writing the analysis and synthesis operations in joint form:

$$\begin{aligned}
f_{t+1}(x, y) &= [f_t(x, y) \otimes HH(x, y)] \otimes \widehat{HH}(x, y) \\
&+ [f_t(x, y) \otimes GH(x, y)] \otimes \widehat{GH}(x, y) \\
&+ [f_t(x, y) \otimes HG(x, y)] \otimes \widehat{HG}(x, y) \\
&+ [f_t(x, y) \otimes GG(x, y)] \otimes \widehat{GG}(x, y)
\end{aligned} \tag{7}$$

where $f(x, y) \otimes GH(x, y)$ is the separable convolution of $f(x, y)$ with $G(x)$ and $H(y)$. Transforming the above equation into Fourier domain, we get:

$$\begin{aligned}
\hat{f}_{t+1}(\omega_x, \omega_y) &= [\hat{f}_t(\omega_x, \omega_y) \cdot \widehat{HH}(\omega_x, \omega_y)] \cdot \widehat{\widehat{HH}}(\omega_x, \omega_y) \\
&+ [\hat{f}_t(\omega_x, \omega_y) \cdot \widehat{GH}(\omega_x, \omega_y)] \cdot \widehat{\widehat{GH}}(\omega_x, \omega_y) \\
&+ [\hat{f}_t(\omega_x, \omega_y) \cdot \widehat{HG}(\omega_x, \omega_y)] \cdot \widehat{\widehat{HG}}(\omega_x, \omega_y) \\
&+ [\hat{f}_t(\omega_x, \omega_y) \cdot \widehat{GG}(\omega_x, \omega_y)] \cdot \widehat{\widehat{GG}}(\omega_x, \omega_y)
\end{aligned} \tag{8}$$

$$\begin{aligned}
\hat{f}_{t+1}(\omega_x, \omega_y) &= \hat{f}_t(\omega_x, \omega_y) \cdot \left\{ [\widehat{HH}(\omega_x, \omega_y) \cdot \widehat{\widehat{HH}}(\omega_x, \omega_y)] \right. \\
&+ [\widehat{GH}(\omega_x, \omega_y) \cdot \widehat{\widehat{GH}}(\omega_x, \omega_y)] + [\widehat{HG}(\omega_x, \omega_y) \cdot \widehat{\widehat{HG}}(\omega_x, \omega_y)] \\
&+ [\widehat{GG}(\omega_x, \omega_y) \cdot \widehat{\widehat{GG}}(\omega_x, \omega_y)] \left. \right\}
\end{aligned} \tag{9}$$

It is obvious from (9) that the perfect reconstruction (i.e., $f_{t+1}(x, y) = f_t(x, y)$) is guaranteed through inverse wavelet transform if,

$$\widehat{HH} \cdot \widehat{\widehat{HH}} + \widehat{GH} \cdot \widehat{\widehat{GH}} + \widehat{HG} \cdot \widehat{\widehat{HG}} + \widehat{GG} \cdot \widehat{\widehat{GG}} = 1 \tag{10}$$

which is usually the case for conventionally used orthogonal and biorthogonal filters (e.g., Haar, Daubechies, and Coiflet).

2.3 Wavelet Diffusion

In this section, we show that the standard discrete wavelet transform can be considered as having the same form as a single discrete step (4) of nonlinear diffusion. Equation (4) can be transformed into Fourier domain:

$$\begin{aligned}
\hat{f}_{t+1}(\omega_x, \omega_y) &= \hat{f}_t(\omega_x, \omega_y) - \omega_x^2 \hat{f}_t(\omega_x, \omega_y) - \omega_y^2 \hat{f}_t(\omega_x, \omega_y) \\
&- i\omega_x \left[\frac{1}{2\pi} \hat{p}(\omega_x, \omega_y) \otimes i\omega_x \hat{f}_t(\omega_x, \omega_y) \right] \\
&- i\omega_y \left[\frac{1}{2\pi} \hat{p}(\omega_x, \omega_y) \otimes i\omega_y \hat{f}_t(\omega_x, \omega_y) \right]
\end{aligned} \tag{11}$$

substituting $1 - \omega_x^2 - \omega_y^2 = \hat{A}_1 \cdot \hat{A}_2$, $i\omega_x = \hat{B}$, $-i\omega_x = (\hat{C} + \hat{D})$, $i\omega_y = \hat{E}$, $-i\omega_y = \hat{F}$, and $\frac{1}{2\pi} \hat{p}(\omega_x, \omega_y) = \hat{p}$, we obtain:

$$\begin{aligned}
\hat{f}_{t+1}(\omega_x, \omega_y) &= \hat{A}_2 \cdot \hat{A}_1 \cdot \hat{f}_t(\omega_x, \omega_y) + \hat{C} \cdot \left[\hat{p} \otimes (\hat{B} \cdot \hat{f}_t(\omega_x, \omega_y)) \right] \\
&+ \hat{D} \cdot \left[\hat{p} \otimes (\hat{B} \cdot \hat{f}_t(\omega_x, \omega_y)) \right] + \hat{F} \cdot \left[\hat{p} \otimes (\hat{E} \cdot \hat{f}_t(\omega_x, \omega_y)) \right]
\end{aligned} \tag{12}$$

Taking the inverse Fourier transform of the above equation:

$$\begin{aligned}
f_{t+1}(x, y) &= f_t(x, y) \otimes \hat{A}_1 \otimes \hat{A}_2 + [p(x, y) \cdot (f_t(x, y) \otimes B)] \otimes C \\
&+ [p(x, y) \cdot (f_t(x, y) \otimes B)] \otimes D + [p(x, y) \cdot (f_t(x, y) \otimes E)] \otimes F
\end{aligned} \tag{13}$$

interestingly, the perfect reconstruction condition (10) is also satisfied in this case by considering the values from the above substitutions,

$$\hat{A}_1 \cdot \hat{A}_2 + \hat{B} \cdot \hat{C} + \hat{B} \cdot \hat{D} + \hat{E} \cdot \hat{F} = 1 \tag{14}$$

Thus equations (12), (13), and (14) have equivalence with (8), (7), and (10),

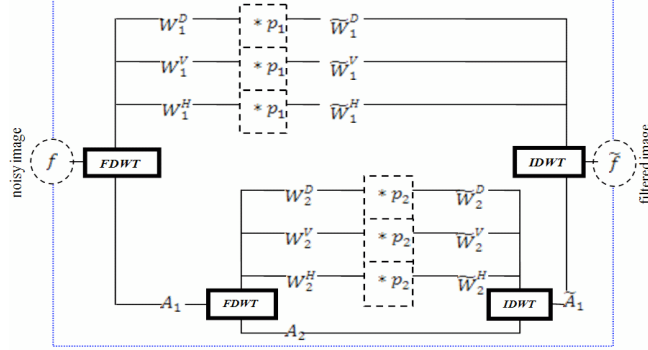


Fig. 1 : Two-level wavelet diffusion

respectively. It is, therefore, deduced from this relationship of the nonlinear diffusion and DWT that wavelet based diffusion works in three steps:

- (i) wavelet decomposition into low-frequency subband (A_j) and high-frequency subbands (W_j^i);
- (ii) regularization of high-frequency coefficients (W_j^i) by multiplication with a function p ; i.e., $\tilde{W}_j^i = p_j * W_j^i$
- (iii) wavelet reconstruction from low-frequency subband (A_j) and regularized high-frequency subbands (\tilde{W}_j^i).

The regularization step multiplies the high-frequency wavelet coefficients by the regularization coefficient $p_j(|\eta_j(x, y)|) = 1 - g(|\eta_j(x, y)|)$ where $g(|\eta_j(x, y)|)$ is the aforementioned nonnegative monotonically decreasing diffusivity function and $|\eta_j(x, y)|$ is the wavelet domain edge estimate for pixel (x, y) at scale j . This process of discrete wavelet diffusion is illustrated in Fig. 1. It must be noted that the wavelet diffusion operation can be considered as an iterative wavelet shrinkage [2] process.

3 Experimental Results and Discussion

In this work, we used the undecimated discrete wavelet transform (UDWT) though DWT is also a possible option. The wavelet filter used was Haar and decomposition level was restricted to four, after initial experimentation with 1 to 5 decomposition levels. Further investigations on the use of decimated discrete wavelet transform, finding the appropriate number of decomposition levels, and the choice of optimal wavelet filter are left for future work. The diffusivity function g employed is the one proposed by Weickert et al. [9]:

$$g(|\eta_j(x, y)|) = \begin{cases} 1, & |\eta_j(x, y)| = 0 \\ 1 - \exp\left[\frac{-3.315}{\left(\frac{|\eta_j(x, y)|}{\lambda}\right)^4}\right], & |\eta_j(x, y)| > 0 \end{cases} \quad (15)$$

which is used to derive the regularization function p_j . The edge estimate $|\eta_j(x, y)|$ is approximated by the modulus of the high-frequency wavelet subbands at each scale j (i.e., each scale has its own edge estimate) and is computed by:

$$|\eta_j(x, y)| = \sqrt{\left(W_j^H(x, y)\right)^2 + \left(W_j^V(x, y)\right)^2 + \left(W_j^D(x, y)\right)^2} \quad (16)$$

A common value of edge threshold λ in (15) was used at all scales. The images were acquired from the “standard” test image collection of [10]. We used a multiscale approach for wavelet diffusion, in comparison to [4], by considering each approximation subband at a particular decomposition level as an independent image and high-frequency wavelet subbands of each level were diffused for noise removal. This lead to fewer iterations (about 5 on average) required for denoising in comparison to the single-scale diffusion of [4] which took nearly 20 or more iterations.

We experimented with different values for λ , the results of which are not shown (due to lack of space) where the noisy images were polluted with additive Gaussian noise of signal-to-noise-ratio (SNR) of 10dB and 15dB. Some visual results for denoising with the proposed wavelet diffusion, Wiener filtering and conventional nonlinear diffusion are presented in Fig. 2. Our results with varying the values of edge threshold λ are in line with the results of [4] and, in general, with the conventional nonlinear diffusion. It indicates that a higher value of λ takes lesser number of iterations to reach the best results but it is prone to oversmooth the image (rather than preserving the edges) in a few iterations. In contrast, a smaller value of λ takes a higher number of iterations to provide the best results but it usually avoids the oversmoothing effect. Moreover, the diffusion process should stop after a fixed number of iterations or it may actually degrade the image quality by oversmoothing effect. However, according to the authors’ knowledge, there is currently no mechanism to determine the best number of diffusion iterations required (either in wavelet based or in conventional non-linear diffusion methods). In addition, we can deduce from these results that a higher value of λ provides good results when the amount of noise is high.

Fig. 2 shows the visual and quantitative results of the proposed wavelet diffusion and compares these with the results of Wiener filtering and conventional nonlinear diffusion in terms of visual quality, SNR, and structural similarity index measure (SSIM) [11]. Apparently, the SNR values of Wiener filtering and nonlinear diffusion for these images are close to the wavelet diffusion results but the visual quality and SSIM values (in the range of 0 to 1) indicate that the results with the proposed method have superior edge preservation, region homogeneity, and similarity to the original image. It was found in [11] that SNR may give flawed measure of image quality as it is not sensitive to the structure (e.g., edges) in an image. It is evident that wavelet diffusion method either provides comparable or favourable results than both Wiener filtering and Weickert nonlinear diffusion. We experimented with a range of parameter values and chose those values, specified in Fig. 2, that produced optimal results with the given set of images.

4 Conclusions

In this paper, we have proposed a generalization of the wavelet diffusion for use with the discrete wavelet transform. We demonstrated that a single discrete step of nonlinear diffusion can be considered as equivalent to the DWT decomposition, high-frequency subbands regularization, and reconstruction. We showed also that the wavelet diffusion can operate in multiscale. Image denoising results of the proposed method compared with Wiener filtering and conventional nonlinear diffusion showed that the technique is promising. Moreover, the method smoothes the noise while preserving the important structures such as edges. Further investigations could be done in the use of DWT instead of UDWT, the number of wavelet decomposition levels, the wavelet bases, number of diffusion iterations, comparison with MZ-DWT based diffusion, and using varying threshold value for different scale. This may require looking for an automatic method of threshold computation.

References

- [1] P. Perona and J. Malik, "Scale Space and Edge Detection Using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629-639, 1990.
- [2] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, vol. 81, pp. 425-455, 1994.
- [3] R. R. Coifman and D. Donoho, "Translation-Invariant Denoising," in *Wavelets in Statistics*, A. Antoine and G. Oppenheim, Eds. New York: Springer, 1995, pp. 125-150.
- [4] A. C.-C. Shih, H.-Y. M. Liao, and C.-S. Lu, "A New Iterated Two-Band Diffusion Equation: Theory and Its Applications," *IEEE Transactions on Image Processing*, 2003.
- [5] Y. Yue, M. M. Croitoru, A. Bidani, J. B. Zwischenberger, and J. W. C. Jr., "Nonlinear Multiscale Wavelet Diffusion for Speckle Suppression and Edge Enhancement in Ultrasound Images," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 297-311, 2006.
- [6] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 710-732, 1992.
- [7] P. Mrazek, J. Weickert, and G. Steidl, "Diffusion-Inspired Shrinkage Functions and Stability Results for Wavelet Denoising," *International Journal of Computer Vision*, vol. 64, pp. 171-186, 2005.
- [8] S. Mallat, *A Wavelet Tour of Signal Processing*. MA, USA: Academic Press, 1998.
- [9] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner, 1998.
- [10] www.imageprocessingplace.com.
- [11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600-612, 2004.

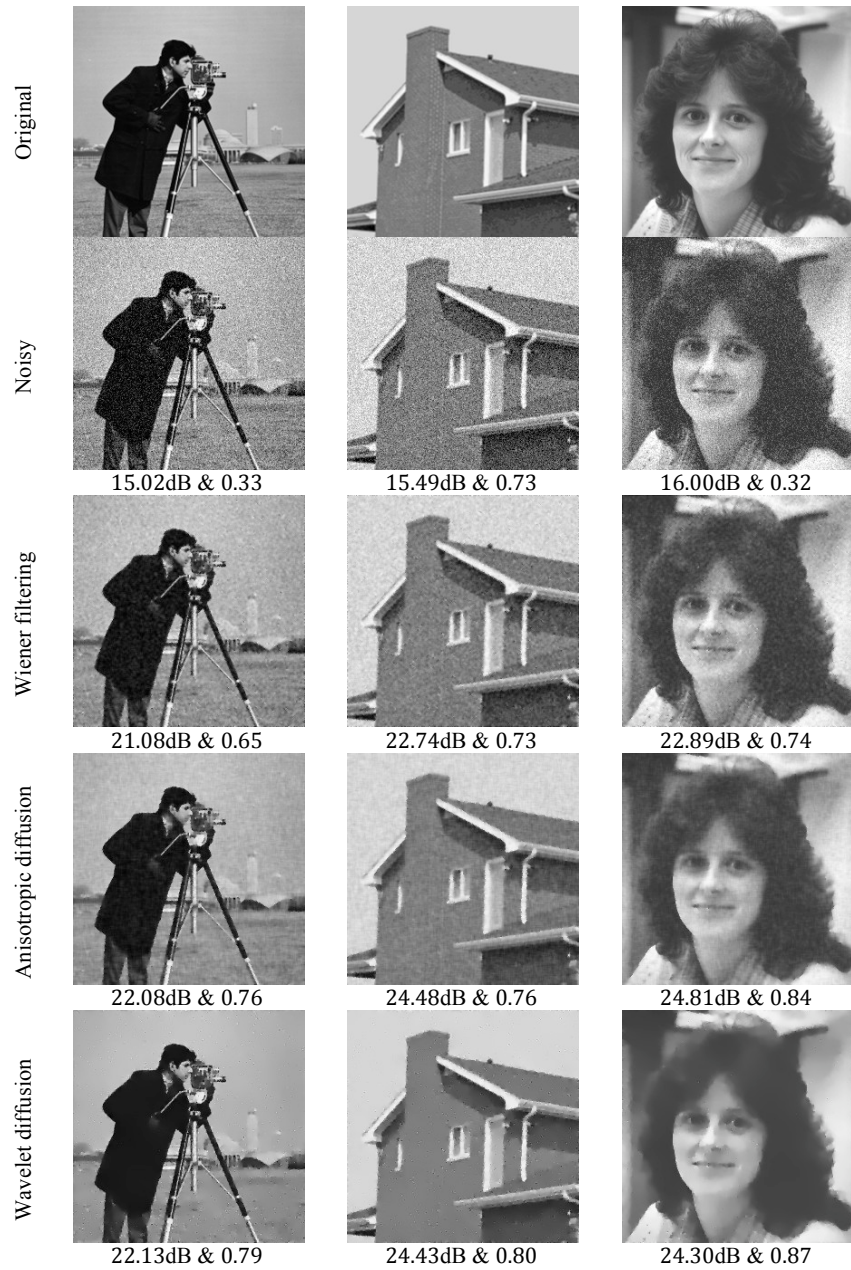


Fig. 2: Visual, SNR, and SSIM results (Anisotropic Diffusion and Wavelet Diffusion: $\lambda = 30$ and iterations = 3; Anisotropic Diffusion: Time-step = 0.5 and Weickert diffusivity function; Wiener filtering: 3x3 window)