A Bayesian formulation of edge-stopping functions in nonlinear diffusion

Aleksandra Pižurica, Iris Vanhamel, Hichem Sahli, Wilfried Philips and Antonis Katartzis

Abstract

We propose a novel, Bayesian formulation of the edge-stopping (diffusivity) function in a nonlinear diffusion scheme in terms of edge probability under a marginal prior on noise-free gradient. This formulation differs from the existing probabilistic diffusion approaches which give stochastic formulations for the conductivity but not for the diffusivity function of the gradient. In particular, we impose a Laplacian prior for the ideal gradient but the proposed formulation is general and can be used with other marginal distributions. The proposed diffusivity function has no free parameters to optimize and it fits well in the cluster of the reference backward-forward diffusivities. We also make links to related works that treat correspondences between nonlinear diffusion and wavelet shrinkage.

Index Terms

Nonlinear diffusion, edge-stopping function, Bayesian diffusivity, Laplacian prior.

I. INTRODUCTION

Nonlinear diffusion [1] has found numerous applications and has attracted research attention in image processing. Let $f(\mathbf{r}) = \{f^{(1)}(\mathbf{r}), f^{(2)}(\mathbf{r}), \ldots, f^{(M)}(\mathbf{r})\}$ denote a M-band image defined on a finite domain Ω, where $\mathbf{r}$ is the position vector on $\mathbf{Z}^2$ defining the coordinates of image pixels. The following system of coupled partial differential equations (PDEs) produces a family of diffused images $u(\mathbf{r}, t)$, which are

A. Pižurica (Aleksandra.Pizurica@telin.UGent.be) and W. Philips (philips@telin.UGent.be) are with the Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium. Tel: +32 9 264 34 12, Fax: +32 9 264 42. I. Vanhamel (iuvanham@etro.vub.ac.be), A. Katartzis (akatarzi@etro.vub.ac.be) and H. Sahli (hsahli@etro.vub.ac.be) are with the Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. Tel: +32 2 629 29 16, Fax: +32 2 629 28 83. Corresponding author A. Pižurica is a postdostoral researcher of the Fund for the Scientific Research (FWO), Flanders, Belgium.

smoothed versions of $f(\mathbf{r})$:

$$\begin{cases} \partial_t u^{(b)}(\mathbf{r}, t) & = \mathrm{div}\left[c(\mathbf{r}, t)\nabla u^{(b)}(\mathbf{r}, t)\right] & \forall b = 1...M \\ u(\mathbf{r}, t = 0) & = f(\mathbf{r}) \\ \partial_{\mathbf{n}} u(\mathbf{r}, t) & = 0 & \text{on } \delta\Omega \end{cases} \tag{1}$$

where ∂_t is the partial derivative over t, $c(\mathbf{r}, t)$ is the so-called *conductance coefficient*, which controls the diffusion process at each position $\mathbf{r}$; t being the continuous scale parameter. $\delta\Omega$ is the image boundary, with $\mathbf{n}$ denoting the normal direction to it. b is the band index.

In a *nonlinear* diffusion the conductance coefficient $c(\mathbf{r}, t)$ varies spatially to encourage intra-region smoothing while preventing the distortion of region boundaries. In the widely adopted formulation by Perona and Malik [1] the conductance coefficient is a function of the gradient magnitude $|\nabla u(\mathbf{r}, t)|$:

$$c(\mathbf{r}, t) = g(|\nabla u(\mathbf{r}, t)|) \tag{2}$$

where $g(x)$ is the so-called *edge stopping* or *diffusivity* function. The diffusivity function is a nonnegative monotonically decreasing function [2]. It ensures that important edges and region boundaries are less blurred than flat low-contrast regions. Common diffusivity functions include those proposed by Perona [1], denoted as *Lorentzian* (g_{Lor}) and *Le Clerc* (g_{Lec}), respectively, the *Tukey bi-weight* function (g_{Tuk}) [3] and the *Weickert* edge stopping function (g_{Wei}) [4]:

$$g_{Lor}(x) = 1/(1 + x^2/k^2)$$

$$g_{Lec}(x) = \exp(-x^2/k^2)$$

$$g_{Tuk}(x) = \begin{cases} (1 - x^2/5k^2)^2 & \text{if } |x| < \sqrt{5}k \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$g_{Wei}(x) = \begin{cases} 1 - \exp(-3.31488k^8/x^8) & \text{if } x \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

The parameter k (contrast parameter) in these functions controls the shape of the diffusivity function, balancing the degrees of inter-region smoothing and edge enhancement in the diffusion process. Black et al [3] showed that various edge-stopping functions in the anisotropic diffusion are closely related to the corresponding error norms and influence functions in the robust statistical estimation framework [5].

The main required property of the edge-stopping functions is that they should have a smaller value for those gradient magnitudes that are more likely to present image edges. In this sense, the edge stopping

functions *implicitly* express the probabilities of the absence of an edge. To our knowledge no attempts were made so far to *explicitly* express the diffusivity function $g(x)$ as the probability that x represents no edge under a suitable marginal prior for the ideal, noise-free image gradient. This problem is the scope of this letter.

In related work, several researchers have recently proposed stochastic formulations of the conductance coefficients in (1) [6], [7], which is usually called "probabilistic" diffusion. These approaches replace the classical diffusivity function of the image gradient by a stochastic conductance coefficient, which is not an explicit function of the image gradient.

In this paper, we return to the classical Perona-Malik formulation of the nonlinear diffusion, where diffusivity is an explicit function of the image gradient and we revisit this formulation within a Bayesian framework. In particular, we impose a prior distribution on the ideal image gradient, and we derive the diffusivity function as the probability that the observed noisy gradient presents no edge of interest. We make a parallel to the related recent developments in wavelet processing [8]. Other researchers have studied connections between discrete nonlinear diffusion schemes and wavelet shrinkage functions [9]–[11]. The correspondence between diffusivity functions and shrinkage estimators in a shift-invariant Haar wavelet representation was recently established in [10], [11]. In this letter, we investigate new relationships between the two domains. This is done by deriving new, Bayesian diffusivity functions based on the statistical image models that were initially developed for wavelet processing.

The letter is organized as follows. Section II presents the proposed Bayesian formulation of the edge-stopping function. We start from the prior specification (Section II-A) and then we formalize the diffusivity in terms of the corresponding edge probability (Section II-B). In Section III, we discuss the shape of the resulting diffusivity function and compare it to other, often used diffusivity functions. The concluding remarks are in Section IV.

II. THE PROPOSED APPROACH

The main idea of our approach is to express the diffusivity function as a probability that the observed gradient presents no edge of interest under a suitable marginal prior distribution for the noise-free gradient.

A. Specifying a suitable prior for the ideal gradient

Noise-free images typically contain large portions of relatively uniform regions that produce negligible gradient values. Thus the noise-free gradient histogram is typically sharply peaked at zero. Sharp edges and textured regions produce some relatively large gradients, building in this way long tails of the gradient

histogram. Hence, candidate priors for noise-free gradients are highly kurtotic, heavy tailed distributions. In the following, we denote by y the noise-free gradient and by x its noisy, observed version.

The *Laplacian* or *double exponential* prior

$$p(y) = (\lambda/2)e^{-\lambda|y|} \tag{4}$$

is often used in wavelet processing [12], [13] because of its analytical simplicity and because in denoising and compression schemes its performance loss compared to more complicated priors like generalized Laplacian [14] is often negligible (see [13] and the references therein). For the noise model

$$x = y + n \tag{5}$$

where n is additive white Gaussian noise $n \sim N(0, \sigma^2)$, the parameter λ is accurately estimated as [15]:

$$\lambda = \left(0.5(\sigma_x^2 - \sigma^2)\right)^{-1/2}. \tag{6}$$

with σ_x^2 denoting the variance of the noisy signal. Since each wavelet subband can be interpreted as a partial derivative of the smoothed image in a given direction [14], marginal statistics of wavelet coefficients directly applies to other smoothed gradient formulations in nonlinear diffusion. In this letter we adopt the prior (4) for which we derive the particular diffusivity function. However, our Bayesian formulation is not restricted to this particular prior.

In nonlinear diffusion schemes, the gradient is often approximated by absolute difference of neighboring pixel values on a 4-connected grid [1]. For color images, the color distances between the neighboring pixels are typically used [16]. Our method does not depend on the particular gradient formula, but note that the estimate of the prior parameter λ in (6) holds only when the noise in the gradient image is approximately additive white Gaussian. In practice, this is often a recommended approximation. Strictly speaking, the property exactly holds when the input image noise is white Gaussian and the gradient is an orthogonal operator (e.g., the orthogonal wavelet transform with the Haar wavelet, which is a moving difference operator).

B. The proposed probabilistic diffusivity function

Similar to some recent wavelet processing schemes [8], [17], we define an edge-element *of interest* as noise-free gradient magnitude that exceeds a specific threshold, T, and we relate this threshold to the noise standard deviation σ as it is explained later in the text. Hence, we formulate two hypotheses: H_0: "*an edge-element of interest is absent*" and H_1: "*an edge-element of interest is present*" as:

$$H_0 : |y| \leq T \quad \text{and} \quad H_1 : |y| > T. \tag{7}$$

We define the diffusivity function of the observed gradient x as

$$g(x) = A(1 - P(H_1|x)) \qquad (8)$$

where A is a normalizing constant. By choosing

$$A = 1/(1 - P(H_1|0)) \qquad (9)$$

we ensure that $g(0) = 1$, because the minimum of $P(H_1|x)$ is at $x = 0$ (see Fig. 1) and thus $1 - P(H_1|x)$ peaks at $x = 0$. The Bayes' rule yields $P(H_1|x) = \mu\eta(x)/[1+\mu\eta(x)]$ where $\mu = P(H_1)/P(H_0)$ presents *prior odds* and $\eta(x) = p(x|H_1)/p(x|H_0)$ is the *likelihood ratio*. Denoting $\eta_0 = p(0|H_1)/p(0|H_0)$, our diffusivity function becomes

$$g(x) = (1 + \mu\eta_0)\frac{1}{1 + \mu\eta(x)}. \qquad (10)$$

Let $\Phi(y)$ denote the cumulative distribution of the standard normal distribution $N(0,1)$ and define $\Psi(a;t) = \Phi(a + t) - \Phi(a)$. For the Laplacian prior (4) we have [8]:

$$\mu = 1/(e^{\lambda T} - 1) \qquad (11)$$

and we can show that (see Appendix)

$$\mu\eta_0 = \frac{1 - \Phi(\sigma\lambda + T)}{\Phi(\sigma\lambda + T) - \Phi(\sigma\lambda)} \qquad (12)$$

and

$$\mu\eta(x) = \frac{r(x/\sigma; \sigma\lambda, T)}{\rho(x/\sigma, \sigma\lambda, T)} \qquad (13)$$

where

$$r(x; \lambda, T) = e^{(x-\lambda)^2/2}\Phi(x - \lambda - T) + e^{(x+\lambda)^2/2}\Phi(-x - \lambda - T), \qquad (14)$$

$$\rho(x; \lambda, T) = e^{(x-\lambda)^2/2}\Psi(\lambda - x; T) + e^{(x+\lambda)^2/2}\Psi(\lambda + x; T). \qquad (15)$$

We define $T = \max(\sigma, \epsilon)$, where ϵ is a small positive constant (we use ϵ=0.2), which ensures the existence of the solution when $\sigma = 0$. When σ decreases, $P(H_1|x)$ has a sharper minimum at $x = 0$ (see Fig. 1) and consequently $g(x)$ has a sharper peak at $x = 0$ meaning no smoothing except at negligible gradient values. In practical diffusion schemes a numerical implementation of the proposed diffusivity function in the form of a look-up-table might be preferred. Such an implementation can significantly accelerate the diffusion process since the calculation of the continuous function at each diffusion step is impractical and time consuming. In a numerical implementation, the diffusivity look-up-table is generated by tabulating the diffusivity function $g(x_q)$ for quantized values of the gradient x_q, using the parameters σ and λ estimated from the input noisy image. Once generated, the look-up-table is used through all the iterations in the diffusion process.
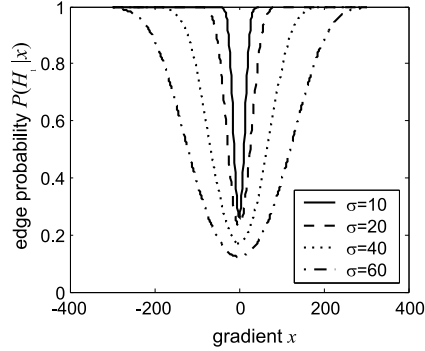
Fig. 1.   Edge probability $P(H_1|x)$ under the Laplacian prior with a fixed parameter λ and for different noise levels σ.

III. SHAPE OF THE PROPOSED FUNCTION

We compared the proposed diffusivity function to the ones given in (3), for which we optimized the diffusion performance by estimating the contrast parameter k according to [3]. We estimate the noise standard deviation σ using a median absolute deviation (MAD) estimator [14], where σ is estimated by the gradient's MAD divided by 0.6745. All the reference functions belong to the class of the so-called *backward-forward* schemes [11], which are designed to smooth weak edges and enhance the strong ones. Our experiments showed that the new probabilistic diffusivity function automatically fits in the cluster of the reference edge-stopping functions with optimized values of the contrast parameter. This is illustrated in Fig. 2 for two different test images. Similar behavior is observed on a series of other tested images. While for the reference functions different procedures exist for estimating the contrast parameter (and in practice often manual tuning), the proposed function leaves no parameters to adjust. The results of the diffusion process using the proposed diffusivity function are illustrated in Fig. 3 in comparison with the standard diffusivity functions. The results of the proposed function are shown in the last two images for the Laplacian prior (i.e., using (12)-(15) in (10)), and for the generalized Laplacian prior (solving (10) numerically). The results demonstrate an improvement over the standard functions and no significant difference in using the generalized Laplacian instead of the Laplacian prior.

It is interesting to notice that a link exists between the proposed Bayesian diffusivity function and the result of [10], [11] that establishes a correspondence between nonlinear diffusion and the wavelet shrinkage. Assume again the noise model (5) and denote a wavelet shrinkage estimator by $\hat{y} = S_y(x)x$. The authors of [10], [11] show that in a shift-invariant wavelet representation (with cycle spinning) using the Haar wavelet, one cycle of the shrinkage estimator amounts to nonlinear diffusion with the
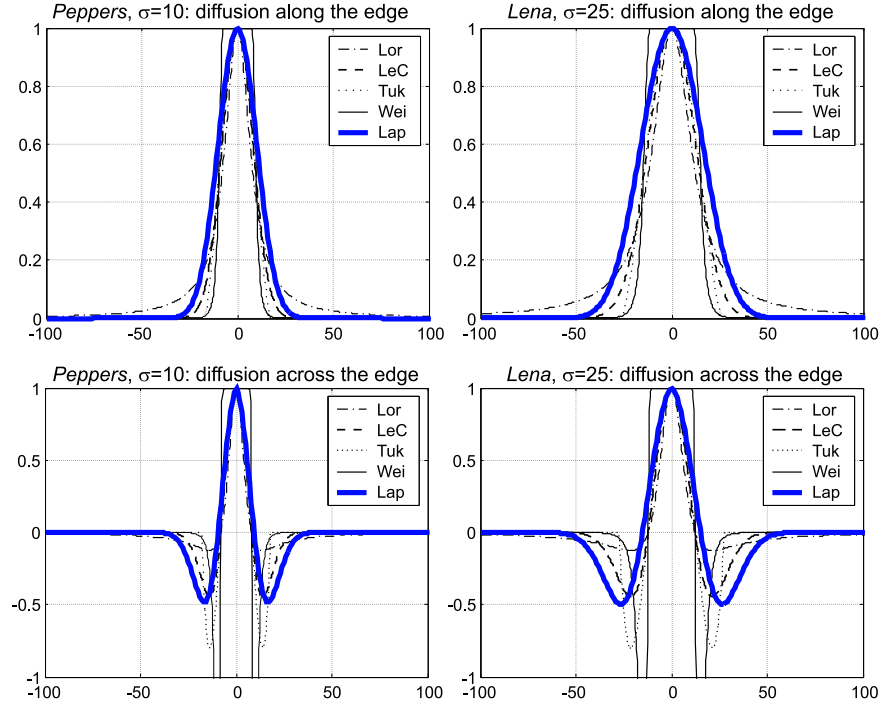
Fig. 2. Top: Diffusivity functions $g(x)$ (diffusion along side the edge). Bottom: functions $[xg(x)]'$ (diffusion across the edge) for two test images. Left: *Peppers*, $\sigma = 10$. Right: *Lena*, $\sigma = 25$. The proposed Bayesian diffusivity under the Laplacian prior (*Lap*) is displayed in comparison to the Lorentzian (*Lor*), LeClerc (*LeC*), Tukey bi-weight (*Tuk*) and Weickert (*Wei*) diffusivities.

edge-stopping function

$$g(|x|) = C\Big(1 - \frac{\sqrt{2}}{x} S_y\Big(\frac{x}{\sqrt{2}}\Big)\Big) \qquad (16)$$

where C is a constant related to the time step in the diffusion process. If we note that the scaling factor $\sqrt{2}$ in the right-hand side of this expression comes from a specific normalization of the gradient (or wavelet coefficient values) and if we define $S_y(x) = P(H_1|x)x$ like in [8], then (16) yields the proposed diffusivity function (8). This means that the proposed Bayesian diffusivity function can be derived from a Bayesian wavelet shrinkage estimator of [8] using the correspondences between the two domains established in [10], [11]. Notice however, that this is only one interpretation of the proposed Bayesian diffusivity function, which is by no means restricted to the analogy with the wavelet shrinkers.

IV. CONCLUSION

We propose a Bayesian formulation of the diffusivity function by imposing a prior on the noise-free gradient. All other probabilistic approaches provide stochastic formulations for the *conductivity* but *not*

Fig. 3.   Results of the diffusion process. Left to right: original image, using Lorentzian ($T_{end} = 27.30$), Le Clerc ($T_{end} = 165$), Weickert ($T_{end} = 165$), Tukey-biweight ($T_{end} = 165$), the proposed function under the Laplacian prior ($T_{end} = 165$) and the proposed function under the generalized Laplacian prior ($T_{end} = 165$).

for the diffusivity function of the gradient. The proposed diffusivity function has no free parameters to optimize and it fits well in the cluster of the reference backward-forward diffusivities. The new function can be derived from a Bayesian wavelet shrinker using the methodology of [10], [11]. The analytical form of the proposed diffusivity function under the Laplacian prior is somewhat complex. The situation is even more complex for the generalized Laplacian. An interesting issue for future investigation is the use of a series expansion of the derived expressions as well as the use of other highly-curtotic priors.

V. APPENDIX A

Define

$$\mathcal{J}(x; a, b, \lambda, \sigma) = \int_a^b e^{-\lambda y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2\sigma^2}} dy. \tag{17}$$

Under the Laplacian prior for noise-free gradient $p(y) = (\lambda/2)e^{-\lambda|y|}$, noise model $x = y + n$, with n statistically independent of y and $n \sim N(0, \sigma^2)$, and for $H_0 : |y| \leq T$, $H_1 : |y| > T$, we have

$$p(x|H_0) = A_0[\mathcal{J}(x; 0, T, \lambda, \sigma) + \mathcal{J}(-x; 0, T, \lambda, \sigma)] \tag{18}$$

$$p(x|H_1) = A_1[\mathcal{J}(x; T, \infty, \lambda, \sigma) + \mathcal{J}(-x; T, \infty, \lambda, \sigma)] \tag{19}$$

with $A_0 = (\lambda/2)e^{\lambda T}/(e^{\lambda T} - 1)$ and $A_1 = (\lambda/2)e^{\lambda T}$.

Let $\phi(x)$ denote zero mean normal density with $\sigma = 1$ and $\Phi(x)$ its cumulative distribution. On a nonnegative interval $[a, b]$ one can verify [13]

$$I(x; a; b, \lambda) = \int_a^b e^{-\lambda y} \phi(x - y) dy =$$
$$\sqrt{2\pi}\phi(x)e^{(x-\lambda)^2/2}[\Phi(b - x + \lambda) - \Phi(a - x + \lambda)]. \tag{20}$$

Using a change of variables $y' = \frac{y}{\sigma}$, which yields $\mathcal{I}(x; a, b, \lambda, \sigma) = \frac{1}{\sigma}I(\frac{x}{\sigma}; a, b, \sigma\lambda)$, and defining

$$r(x; \lambda, T) = e^{(x-\lambda)^2/2}\Phi(x - \lambda - T) + e^{(x+\lambda)^2/2}\Phi(-x - \lambda - T)$$

$$\rho(x; \lambda, T) = e^{(x-\lambda)^2/2}\Psi(\lambda - x; T) + e^{(x+\lambda)^2/2}\Psi(\lambda + x; T)$$

where $\Psi(a; T) = \Phi(a + T) - \Phi(a)$, we have

$$p(x|H_0) = \frac{A_0}{\sigma} \phi(\frac{x}{\sigma}) \rho(\frac{x}{\sigma}, \sigma\lambda, T) \tag{21}$$

$$p(x|H_1) = \frac{A_1}{\sigma} \phi(\frac{x}{\sigma}) r(\frac{x}{\sigma}, \sigma\lambda, T) \tag{22}$$

and the ratio $\eta(x) = p(x|H_1)/p(x|H_0)$ is thus

$$\eta(x) = (e^{\lambda T} - 1) \frac{r(\frac{x}{\sigma}, \sigma\lambda, T)}{\rho(\frac{x}{\sigma}, \sigma\lambda, T)}. \tag{23}$$

For $x = 0$, we have $r(0; \lambda, T) = 2e^{\frac{\lambda^2}{2}}[1 - \Phi(\lambda + T)]$ and $\rho(0; \lambda, T) = 2e^{\frac{\lambda^2}{2}}\Psi(\lambda; T)$, and thus $\eta_0 = p(0|H_1)/p(0|H_0)$ becomes

$$\eta_0 = (e^{\lambda T} - 1) \frac{1 - \Phi(\sigma\lambda + T)}{\Phi(\sigma\lambda + T) - \Phi(\sigma\lambda)}. \tag{24}$$

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] S. Z. Li, "Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 20, no. 9, pp. 916–932, 1998.

[3] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[4] J. Weickert, *Anisotropic diffusion in image processing*, ser. ECMI Series. Stuttgart, Germany: Teubner-Verlag, 1998.

[5] A. Ben Hamza and H. Krim, "Image denoising: A nonlinear robust statistical approach," *IEEE Trans. on Signal Processing*, vol. 49, no. 12, pp. 3045–3054, 2004.

[6] S. Arridge and A. Simmons, "Multi-spectral probabilistic diffusion using bayesian classification," in *Proc. Scale-Space '97*, 1997, pp. 224–235.

[7] Y. Bao and H. Krim, "Smart nonlinear diffusion: A probabilistic approach," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26(1), pp. 63–72, 2004.

[8] A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Processing (In press)*.

[9] Y. Bao and H. Krim, "Towards bridging scale-space and multiscale frame analyses," in *Wavelets in Signal and Image Analysis: From Theory to Practice, Computational Imaging and Vision*, A. Petrosian and F. Meyer, Eds., vol. 19. Kluwer Pub., 2001.

[10] P. Mrázek, J. Weickert, and G. Steidl, "Correspondences between wavelet shrinkage and nonlinear diffusion," in *4th International Conference on Scale Space Methods in Computer Vision (Scale-Space 2003)*, ser. Lecture Notes in Computer Science, L. Griffin and M. Lillholm, Eds., vol. 2695. Isle of Skye-UK: Springer, 2003, pp. 1001–116.

[11] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Weber, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004.

[12] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, pp. 909–919, Apr. 1999.

[13] M. Hansen and B. Yu, "Wavelet thresholding via MDL for natural images," *IEEE Trans. Inform. Theory*, vol. 46, no. 5, pp. 1778–1788, Aug. 2000.

[14] S. Mallat, *A wavelet tour of signal processing*. Academic Press, London, 1998.

[15] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, Lausanne, Switzerland, 1996.

[16] D. Androutsos, K. Plataniotis, and A. Venetsanopoulos, "Distance measures for color image retrieval," in *ICIP '98*, Chicago, Illinois - USA, 1998, pp. 770–774.

[17] M. Jansen and A. Bultheel, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 629–639, 2001.