IMAGE DENOISING IN NONLINEAR SCALE-SPACES: AUTOMATIC SCALE SELECTION VIA CROSS-VALIDATION

George Papandreou and Petros Maragos

National Technical University of Athens, School of ECE, 15773, Athens, Greece
E-mail: {gpapan,maragos}@cs.ntua.gr

ABSTRACT

Multiscale, i.e. *scale-space* image analysis is a powerful framework for many image processing tasks. A fundamental issue with such scale-space techniques is the automatic selection of the most salient scale for a particular application. This paper considers optimal scale selection when nonlinear diffusion and morphological scale-spaces are utilized for image denoising. The problem is studied from a statistical model selection viewpoint and cross-validation techniques are utilized to address it in a principled way. The proposed novel algorithms do not require knowledge of the noise variance, have acceptable computational cost and are readily integrated with a wide class of scale-space inducing processes which require setting of a scale parameter. Our experiments show that this methodology leads to robust algorithms, which outperform existing scale-selection techniques for a wide range of noise types and noise levels.

1. INTRODUCTION

A primary concern for many applications of image analysis and computer vision is the presence of noise. Most algorithms for image segmentation, surveillance or medical image analysis require effective noise suppression to produce reliable results. Therefore an image denoising module is often present in the pre-processing stage of many practical image analysis systems.

Nonlinear scale-spaces, induced either by nonlinear diffusion processes or morphological operators, are particularly effective at suppressing noise while preserving important image features e.g. edges and are widely used for image denoising. Nonlinear diffusion *Partial Differential Equations (PDEs)* which generate general-purpose denoising methods are generalizations of the homogenous heat equation. Representative examples are the Perona-Malik nonlinear edge-preserving diffusion equation [1], with its mathematically well-posed modification [2], and the anisotropic diffusion equation [3]. Variants of these denoising PDEs for specific applications have also been proposed (e.g. [4]). In the experiments reported in Sec. 4 we have used the diffusion PDE of [2] (from now on, x denotes a 2D position vector on the image plane):

$$\frac{\partial u(x,t)}{\partial t} = \operatorname{div}\big(g(\|\nabla u_\sigma\|)\nabla u\big), \tag{1}$$

with initial condition $u(x,0) = y(x)$, which creates a scale-space $\{f_t : t \geq 0\}$, with $f_t(x) \triangleq u(x,t)$. In Eq. (1) the diffusivity $g : [0,+\infty) \to R^+$ is a decreasing function (with $g(0) = 1$, $g(r) \to 0$ while $r \to +\infty$) which favors intraregion over interregion smoothing and u_σ denotes convolution of u with a Gaussian kernel of standard deviation σ.

A large class of morphological filters also generates nonlinear scale-spaces with good denoising properties. The simplest example is Minkowski openings and closings, which are serial compositions of flat erosions and dilations by disks. Cascading open-closings at increasing scales yields the *alternating sequential filter (ASF)* [5]. Further, the Minkowski open-closings in an ASF can be replaced by other types of lattice-theoretic open-closings, such as the *reconstruction* filters [6]. Most of the above morphological scale-spaces can also be implemented with PDEs, which has many advantages over implementation with digital filters [7]. However, reconstruction openings and closings are *not* self-dual operators, treating the image and its background asymmetrically. A newer morphological operator type that unifies both of them and possesses self-duality is the *leveling* [8]. Levelings are nonlinear object-oriented filters that simplify a reference image f by simultaneously locally expanding or shrinking an initial seed image, called the marker m, and globally constraining the marker evolution by the reference image. Let $u(x,t)$ represent the evolutions of m. Then u is a weak solution of the PDE [8, 9] $\partial u(x,t)/\partial t = -\operatorname{sgn}(u-f)\|\nabla u\|$, with initial condition $u(x,0) = m(x)$. This PDE has a non-trivial steady-state [9] $\Lambda(m|f) \triangleq \lim_{t\to\infty} u(x,t)$, which is the leveling of f with respect to m. Now, let us consider various markers $m_t, t = 1, 2, \ldots$, that are related to some increasing scale parameter t and construct the levelings $f_t = \Lambda(m_t|f_{t-1})$, $t > 1$, with $f_0 = y$. The signals $\{f_t : t \geq 0\}$ constitute a hierarchy of *multiscale levelings* of the initial image y possessing the causality property that f_j is a leveling of f_i for $j > i$. A way to construct such multiscale levelings is to use a sequence of multiscale markers obtained by sampling a Gaussian scale-space of y.

Any of the above multiscale denoising methods starts from a noisy initial image y and generates a sequence of scale-space snapshots $\{f_t : t \geq 0\}$, with $f_0 = y$, which depend on an increasing scale parameter t. Thus, a cental issue is how to optimally select the scale at which to stop smoothing the image. At the one extreme, at a very small scale the noise has not yet been suppressed sufficiently and the estimate demostrates high variance. At the other extreme, at a very large scale, not only the noise but also the details of the image have been eliminated – the estimate is highly biased. However, most scale-space representations do not exhibit an inherent stopping point. For example, the steady-state solution of nonlinear diffusion is typically the trivial constant-valued image. Even if the diffusivity g in Eq. (1) is selected in a way that the steady-state f_∞ is non-trivial [10, 11], there is still no guarantee that f_∞ is optimal for the denoising problem in any reason-

Our work has been supported by the European Network of Excellence MUSCLE, the Greek research program 'Pythagoras', the NTUA research program 'Protagoras' and the Greek State Scholarships Foundation.

able way. One clearly needs an automatic method to successfully resolve this bias–variance tradeoff, stopping the denoising procedure after the noise has been adequately suppressed but before the image has been oversmoothed. The need for explicit scale selection mechanisms has also been stressed out by other researchers in similar contexts (see e.g. the work of Lindeberg in [12]).

In our work, we consider choosing the optimal scale for denoising as a statistical model selection problem. We suggest that a desirable scale-selection strategy should choose the scale-space snapshot that minimizes the expected "distance" to the (generally unknown) noise-free image, under a given problem-specific loss function. We propose two novel algorithms based on statistical cross-validation techniques that approximate this ideal optimality criterion in a principled way. These algorithms prove to be particularly effective and robust in practice. Our approach is on the same spirit with [13], where similar cross-validation techniques are utilized for selecting the optimal threshold for denoising by wavelet thresholding. This paper extends our recent work in [14], which was confined to scale selection in diffusion scale-spaces, by studying the performance of the proposed cross-validation scale selection algorithms in denoising schemes utilizing a wider class of nonlinear scale-spaces, including morphologically-induced ones.

2. PROBLEM FORMULATION AND PREVIOUS WORK

Available is a "noisy" grayscale image y consisting of $M = M_x M_y$ pixels. Consider a lexicographic ordering of the pixels and denote with y_i the intensity at the i-th pixel, with $1 \leq i \leq M$. We assume that y is a realization of a random process Y, modeled by:

$$Y_i = f^*(x_i) + N_i, \quad i = 1, \ldots, M, \tag{2}$$

where f^* denotes the usually unknown "clean" image and N is a zero-mean noise process, with independent elements N_i of variance σ_i^2. Starting from the noisy image $y(x)$ as initial condition f_0, the scale-space members $\{f_t : t \geq 0\}$, indexed by an increasing scale parameter t, are candidates for approximating f^*. We seek for the optimal denoised version of our degraded image among these scale-space snapshots f_t.

For that purpose an optimality criterion needs to be established. Suppose that a new instance $y^{new} = f^* + n^{new}$ is generated from Y, with n_i^{new} independent from n_i but identically distributed with it. We define the *(in-sample) prediction error* made by model f_t as:

$$\mathrm{PE}(t) \equiv \mathrm{PE}(f_t) = E_N\{L(y^{new}, f_t)\}, \tag{3}$$

taking expectations over the noise process $\{n_i^{new}\}$, where L is a pixel-normalized loss function which penalizes the deviation between y^{new} and f_t. Typical examples of loss functions are the $L_p(x,y) = \left(\frac{1}{M}\sum_{i=1}^{M}|x_i - y_i|^p\right)^{1/p}$ for $p \in \{1, 2\}$, although other choices might be more appropriate for certain applications. The *optimal stopping time* t^* can then reasonably be defined as the scale that $\mathrm{PE}(t)$ attains its minimum, i.e. $t^* = \mathrm{argmin}_{t\geq 0}\,\mathrm{PE}(t)$.

Another quantity useful in our analysis is the *model error*, defined as $\mathrm{ME}(t) \equiv \mathrm{ME}(f_t) = L(f^*, f_t)$. In the case that the noise power tends to zero, $\mathrm{ME}(t)$ and $\mathrm{PE}(t)$ coincide. Otherwise the added uncertainty due to noise leads to $\mathrm{PE}(t) > \mathrm{ME}(t)$. For example, for square loss and zero-mean i.i.d. errors of variance σ^2, one can easily show that $\mathrm{PE}(t) = \mathrm{ME}(t) + \sigma^2$. The utility of $\mathrm{ME}(t)$ is limited in practice, since f^* is usually unknown. However one can use $\mathrm{ME}(t)$ and the scale $t^*_{\mathrm{ME}} = \mathrm{argmin}_{t\geq 0}\,\mathrm{ME}(t)$ where it attains its minimum as reference when f^* is available, as is the case with the experiments of Sec. 4, where we artificially add noise to images and hence know f^*.

The *extra-sample prediction error*, defined as

$$\mathrm{PE}_+(t) \equiv \mathrm{PE}_+(f_t) = E_{N,X}\{L(y^{new}, f_t)\}, \tag{4}$$

differs from $\mathrm{PE}(t)$ because it treats the position X of pixels in the newly generated image y^{new} as random variables [15, 16]. This means that for $\mathrm{PE}_+(t)$ we might need to compute the image intensity at a point not present in the training set $\{x_i : 1 \leq i \leq M\}$ (cf superresolution). The additional uncertainty introduced in this case means that typically $\mathrm{PE}_+(t) \geq \mathrm{PE}(t)$. In image processing terminology, $\mathrm{PE}_+(t)$ measures not only the uncertainty due to noise and model error, as $\mathrm{PE}(t)$ does, but an interpolation error, as well. However, in our application we are not interested in $\mathrm{PE}(t)$ per se but in the scale t^* that it attains its minimum. Assuming that the interpolation error terms in $\mathrm{PE}_+(t)$ for the different models $\{f_t : t \geq 0\}$ effectively cancel out, it is plausible to take $t^* \approx t^*_+$, where $t^*_+ = \mathrm{argmin}_{t\geq 0}\,\mathrm{PE}_+(t)$ is the scale where $\mathrm{PE}_+(t)$ is minimized. The gain from this assumption is that $\mathrm{PE}_+(t)$ can be estimated directly from the noisy image y using cross-validation techniques, as we will see in Sec. 3.

We briefly review next a number of other approaches to scale selection, mainly from the scale-space related literature. In [17] Weickert selects the scale t^*_{snr} which satisfies the relation $\frac{\mathrm{var}(f_{t^*_{\mathrm{snr}}})}{\mathrm{var}(f_0)} = \frac{1}{1+1/\mathrm{snr}}$, assuming that the signal-to-noise ratio snr is known and that the diffusion filter is so effective, that $f_{t^*_{\mathrm{snr}}}$ is a good approximation of f^*. Mrazek in [18] proposes a decorrelation criterion, selecting the scale t^*_{dec} which minimizes the correlation coefficient $t^*_{\mathrm{dec}} = \mathrm{argmin}_{t\geq 0}\,\frac{\mathrm{cov}(f_0 - f_t, f_t)}{\sqrt{\mathrm{var}(f_0 - f_t)\mathrm{var}(f_t)}}$, considering $f_0 - f_t$ as "noise" and $f_{t^*_{\mathrm{dec}}} \approx f^*$. However the decorrelation criterion does not seem to be connected with any other criterion of filtering quality, which was also noted in [18]. Finally, Solo in [19], assuming i.i.d. gaussian noise of known variance and quadratic penalty, derives a SURE-based criterion for selecting the scale that minimizes $\mathrm{PE}(t)$ under L_2 loss. From the approaches just described, none can handle arbitrary loss functions L. Moreover, the first two of them lack statistical foundations. The cross-validation algorithms we discuss next try to overcome these shortcomings.

3. SCALE SELECTION BY CROSS-VALIDATION

Cross-validation methods attempt to directly estimate the extra-sample prediction error $\mathrm{PE}_+(t)$ of Eq. (4) in a non-parametric, data-driven way [20]. Cross-validatory techniques for model selection are very general, in the sense that they can be used with any loss function L or nonlinear model generator, unlike other approaches to model selection, such as those based on the *Akaike Information Criterion*, *Mallows' Cp statistic* or the *Bayesian Information Criterion*, whose applicability is typically restricted to linear model settings with quadratic loss and known noise variance [21]. Therefore cross-validation is particularly appropriate for scale selection in the context of denoising in nonlinear scale-spaces under arbitrary loss function. Note, however, that for cross-validation techniques to work well in our problem, it is important that the noise elements N_i at different pixels are uncorrelated. For example, it has been shown in [22] (in the context of kernel smoothing) that if noise at neighboring pixels is positively (negative) correlated, then models selected by unadapted cross-validation tend to overfit (resp. underfit) the data (cf. [13]).

In order to estimate $\mathrm{PE}_+(t)$ by cross-validation, we need to properly resample the noisy data $\mathcal{D} = \{(x_i, y_i) : 1 \le i \le M\}$ which form our training set. Let $\mathcal{T}_1 \subset \mathcal{D}$ contain some data withheld from the training set. Then the remaining data from $\mathcal{D}$ can be used to build a nonlinear diffusion scale-space denoted by $f_t^{-\mathcal{T}_1}$. The predictive power of $f_t^{-\mathcal{T}_1}$ can be assessed on $\mathcal{T}_1$ by $\widehat{\mathrm{PE}}_+(f_t^{-\mathcal{T}_1}) = L(y^{\mathcal{T}_1}, f_t^{-\mathcal{T}_1})$, since $\mathcal{T}_1$ is *independent* from the data used to build the model. Repeating K times for different subsets $\mathcal{T}_1, \ldots, \mathcal{T}_K$ of roughly the same size and averaging yields:

$$\widehat{\mathrm{PE}}_+(t) = \mathrm{PE}^{CV}(t) = \frac{1}{K} \sum_{k=1}^{K} \widehat{\mathrm{PE}}_+(f_t^{-\mathcal{T}_k}) \qquad (5)$$

In order to derive practical algorithms for our application, we need to specify the $\{\mathcal{T}_i : 1 \le i \le K\}$. This issue has attracted considerable attention in the literature (see e.g. [23]). For some model selection problems, like choosing the regularization parameter of smoothing splines, *leave-one-out* cross-validation (where $K = M$ and $\mathcal{T}_i$ only contains (x_i, y_i)) can be approximated analytically, leading to fast computations [21]. However leave-one-out cross-validation in our case requires building nonlinear diffusion scale-spaces of M images of size $(M-1)$-pixels each, something unacceptably expensive. Hence we propose two other data resampling configurations, called *quadruple-cv* and *double-cv* from now on. Fig. 1 depicts for each of them the pixels $\mathcal{D} - \mathcal{T}_i$ used to build the scale-space model (in black) and the pixels $\mathcal{T}_i$ used to estimate its prediction error (in white) for the case $i = 1$.

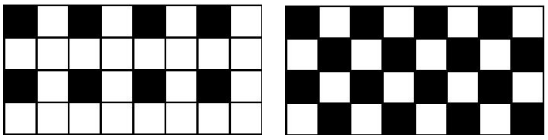

Fig. 1. Resampling configurations. $\mathcal{T}_1$ is depicted in white. *Left:* Quadruple-CV. *Right:* Double-CV.

In *quadruple-cv*, by selecting four different values {(0,0), (0,1), (1,0), (1,1)} for the shift vector (s_i, s_j), we get four subsampled by a factor of two in each direction versions of the noisy image y. Each consists of roughly $M/4$ pixels with coordinates $\{(2i + s_i, 2j + s_j) : 0 \le i < M_x/2, 0 \le j < M_y/2\}$. We then build the scale-space of each of these subsampled images, appropriately scaling the PDEs or the structuring elements of the morphological filters. For example, if the PDE (1) is utilized, we must set $\tau' = \tau/4$ (for the time-step), $g'(\cdot) = g(\cdot/2)$ and $\sigma' = \sigma/2$. We then get $\mathrm{PE}^{CV}_{\mathrm{quadruple-cv}}(t)$ from Eq. (5), computing each of the four terms $\{\widehat{\mathrm{PE}}_+(f_t^{-\mathcal{T}_k}) : 1 \le k \le 4\}$ as follows: We take the corresponding snapshot of the auxiliary scale-space, with dimensions $M_x/2$ by $M_y/2$ and interpolate from it the values at the remaining $3M/4$ pixels $p_i \in \mathcal{T}_k$ (by bilinear interpolation in our implementation). We then penalize (by means of L) the discrepancy between the interpolated value and the initial noisy value y_i at the same pixel p_i and average over the $3M/4$ pixels of $\mathcal{T}_k$ to get $\widehat{\mathrm{PE}}_+(f_t^{-\mathcal{T}_k})$. The computational overhead roughly equals the cost of generating the original scale-space, since 4 auxiliary scale-spaces of size $M/4$ pixels each need to be built.

The main difference in the case of *double-cv* is that the pixels building each of the auxiliary scale-spaces are not located on a rectangular lattice any more (see Fig. 1, right). Therefore, it is convenient first to interpolate the values at the $M/2$ "white" pixels of $\mathcal{T}_k$ from the values at the remaining $M/2$ "black" pixels of the noisy image y_i and then build the auxiliary scale-spaces (no rescaling is thus needed). The overhead of the procedure is twice the cost of the standard scale-space, since two auxiliary full-sized scale-spaces evolve in parallel with the main one.

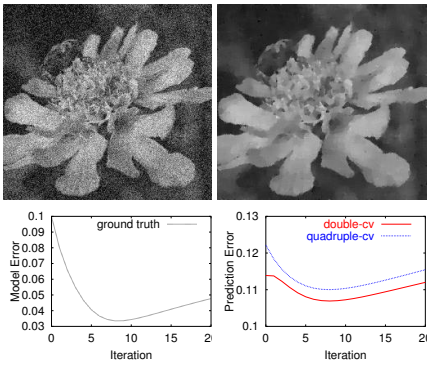

Fig. 2. Image denoising example. *Up:* The noisy image and its denoised version at an automatically selected scale. *Down:* Model Error $\mathrm{ME}(t)$ (ground truth) and Prediction Error $\mathrm{PE}_+(t)$ (estimated by each of the two cross-validation algorithms).

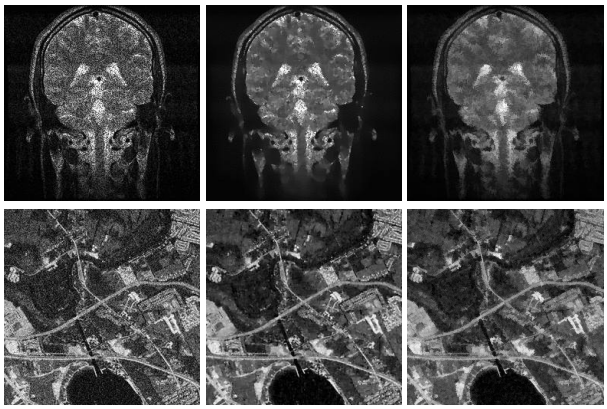

Fig. 3. A noisy image (*left*), as denoised by diffusion (*center*) and leveling (*right*) scale-spaces at automatically selected scales. *Up:* MRI scan, speckle noise. *Down:* Aerial image, gaussian noise.

4. EXPERIMENTS AND COMPARISONS

An example of image denoising with automatic scale selection by cross-validation can be seen in Fig. 2. The first row shows the noisy image and its denoised version at scale $t^*_{\mathrm{double-cv}}$ determined by the *double-cv* cross-validation algorithm. The corresponding plots of the $\mathrm{ME}(t)$ (ground truth) and $\mathrm{PE}_+(t)$ (as estimated by the two cross-validation algorithms) can be seen at the second row. Notice that, as we discussed in Sec. 2, $\mathrm{ME}(t)$ is smaller than $\mathrm{PE}_+(t)$. However, both quantities attain their minimum at roughly the same scale (after about 8 iterations). Further denoising examples by diffusion and leveling scale-spaces, one of an MRI scan and one of an aerial image, can be seen in Fig. 3.

In order to systematically assess the performance of the proposed algorithms and compare them with other techniques, we

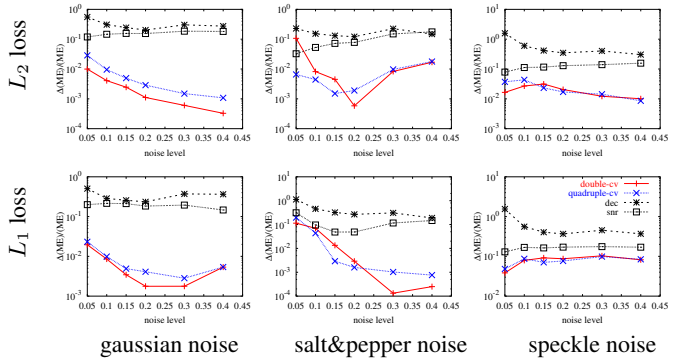

Fig. 4. Relative increase in model error $\Delta(\mathrm{ME})/\mathrm{ME}$ as a function of the noise level nl for the scale selection algorithms under consideration. Average results computed over 39 images. (Log-scale on the y-axis, legend on the bottom-right figure.)

run a series of denoising experiments on a dataset of 39 natural grayscale images (the *kodak*, *aerial* and *misc1* collections available from `http://www.cipr.rpi.edu/resource/stills/`), corrupted by artificial noise so that the ground truth f^* would be available. Apart from the two cross-validation algorithms we propose (*double-cv* and *quadruple-cv*), we have also implemented the snr-based method of Weickert (with ground-truth snr value), and the decorrelation method of Mrazek, in the sequel dubbed *snr* and *dec*, respectively. The scale-space used in the experiments reported here was generated by the diffusion PDE (1), with $g(r) = 1/[1 + (r/\lambda)^2]$ [1], $\lambda = 0.01$ (the intensity values of the images were between 0 and 1) and the AOS scheme [3].

We experimented with three different noise types $nt \in \{$ gaussian, salt&pepper, speckle$\}$. The degraded images were respectively generated by $y_i = f^*(x_i) + \epsilon_i$ (gaussian), $y_i = (1 + \epsilon_i) f^*(x_i)$ (speckle), and $y_i = f^*(x_i)$, with probability $1 - p$ and $y_i = 0$ or 1 with probability $p/2$ each (salt&pepper). In the gaussian and speckle cases, the ϵ_i were i.i.d. sampled from $N(0, \sigma^2)$. We conducted tests for varying noise levels $nl \in \{0.05, 0.1, 0.15, 0.2, 0.3, 0.4\}$, where $nl = \sigma$ in the case of gaussian or speckle noise and $nl = p$ in the case of salt&pepper noise and for two different choices of the loss function $L \in \{L_1, L_2\}$. To assess the performance of the algorithms under consideration for each of the $2 \cdot 3 \cdot 6 = 36$ combinations of (L, nt, nl) we run experiments on all 39 images in the database and averaged the *relative increase in model error*, defined as $\frac{\Delta(\mathrm{ME})}{\mathrm{ME}} = \frac{\mathrm{ME}(t^*_{\mathrm{alg}}) - \mathrm{ME}(t^*_{\mathrm{ME}})}{\mathrm{ME}(t^*_{\mathrm{ME}})}$.

In Fig. 4 we present average benchmark scores for the $\frac{\Delta(\mathrm{ME})}{\mathrm{ME}}$. Notice that *double-cv* and *quadruple-cv* give results that sometimes are one order of magnitude better than the results given by the *snr* and *dec* algorithms. This is particularly true in the case of gaussian noise, where the $\frac{\Delta(\mathrm{ME})}{\mathrm{ME}}$ of the two cross-validation algorithms is almost always less than 1%. The robustness of both cross-validation algorithms, irrespectively of the noise type or the utilized loss function is also noteworthy. Among the other two algorithms, *snr* seems to perform better than *dec*. The decorrelation algorithm *dec* performed rather poorly in our experiments. Note that the experiments in [18] also imply that *dec* doesn't perform well with the nonlinear diffusion PDE (1).

These experimental results demonstrate that cross-validation is very efficient for scale selection in nonlinear scale-spaces, giving good results for a range of noise types and loss functions.

5. REFERENCES

[1] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE PAMI*, vol. 12, no. 7, pp. 629–639, 1990.

[2] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM J. Num. Anal.*, vol. 29, no. 1, pp. 182–193, 1992.

[3] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner-Verlag, 1998.

[4] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. on Im. Pr.*, vol. 11, pp. 1260–1270, 2002.

[5] J. Serra, Ed., *Image Analysis and Mathematical Morphology*, vol. 2, Acad. Press, 1998.

[6] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Trans. on Im. Pr.*, vol. 4, pp. 1153–1160, 1995.

[7] P. Maragos, "PDEs for morphological scale-spaces and eikonal applications," in *The Handbook of Image And Video Processing*, A.C. Bovik, Ed. Elsevier, 2nd edition, 2005.

[8] F. Meyer and P. Maragos, "Nonlinear scale-space representation with morphological levelings," *J. Visual Communic. and Image Representation*, vol. 11, 2000.

[9] P. Maragos, "Algebraic and PDE approaches for lattice scale-spaces with global constraints," *IJCV*, vol. 52, no. 2-3, pp. 121–137, 2003.

[10] M. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. on Im. Pr.*, vol. 7, no. 3, pp. 421–432, 1998.

[11] G. Gilboa, Y. Y. Zeevi, and N. Sochen, "Image enhancement, segmentation and denoising by time dependent nonlinear diffusion processes," in *Proc. Int'l Conf. on Im. Pr.*, 2001.

[12] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, vol. 30, no. 2, pp. 79–116, 1998.

[13] G. P. Nason, "Wavelet shrinkage using cross-validation," *J. R. Stat. Soc. B*, vol. 58, no. 2, pp. 463–479, 1996.

[14] G. Papandreou and P. Maragos, "A cross-validatory statistical approach to scale selection for image denoising by nonlinear diffusion," in *Proc. Int'l Conf. on CVPR*, 2005.

[15] B. Efron, "How biased is the apparent error rate of a prediction rule?," *J. Am. Stat. Assoc.*, vol. 81, pp. 461–470, 1986.

[16] L. Breiman, "The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error," *J. Am. Stat. Assoc.*, vol. 87, pp. 738–754, 1992.

[17] J. Weickert, "Coherence-enhancing diffusion of colour images," *Im. and Vis. Comp.*, vol. 17, pp. 201–212, 1999.

[18] P. Mrazek and M. Navara, "Selection of optimal stopping time for nonlinear diffusion filtering," *IJCV*, vol. 52, no. 2/3, pp. 189–203, 2003.

[19] V. Solo, "Automatic stopping criterion for anisotropic diffusion," in *Proc. ICASSP*, 2001, vol. 6, pp. 3929–3932.

[20] M. Stone, "Cross-validatory choice and assessment of statistical predictions (with discussion)," *J. R. Stat. Soc. B*, vol. 36, pp. 111–147, 1974.

[21] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, 2001.

[22] N. S. Altman, "Kernel smoothing of data with correlated errors," *J. Am. Stat. Assoc.*, vol. 85, pp. 749–759, 1990.

[23] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int'l Joint Conf. on Artif. Intel.*, 1995, pp. 1137–1143.