

A Cross-Validatory Statistical Approach to Scale Selection for Image Denoising by Nonlinear Diffusion

George Papandreou and Petros Maragos

National Technical University of Athens
School of Electrical and Computer Engineering
Zografou 15773, Athens, Greece

E-mail: {gpapan, maragos}@cs.ntua.gr

Abstract

Scale-spaces induced by diffusion processes play an important role in many computer vision tasks. Automatically selecting the most appropriate scale for a particular problem is a central issue for the practical applicability of such scale-space techniques. This paper concentrates on automatic scale selection when nonlinear diffusion scale-spaces are utilized for image denoising. The problem is studied in a statistical model selection framework 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 and have acceptable computational cost. Extensive experiments on natural images show that the proposed methodology leads to robust algorithms, which outperform existing 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. Noise is unavoidably introduced during the image acquisition process for various reasons, including thermal effects, camera defects and the quantum nature of light. The transmission of an image over some communication medium may also further deteriorate its quality. Therefore image denoising techniques are often utilized in the pre-processing stage of many practical image analysis systems. Various computer vision algorithms for multiscale image segmentation or shape recognition require effective noise suppression to produce reliable results. In some applications, like extracting information from SAR images or processing PET medical images, the effect of noise is so strong, that the role of a denoising module is indispensable.

An important category of image denoising methods utilize various PDE-based diffusion processes. Evolving the

noisy image under such laws yields a multiscale analysis of the noisy image. An estimate of the “clean” image is selected among the scale-space snapshots by choosing an appropriate value for the scale parameter, which in this paper will coincide with the evolution time of the diffusion. The class of diffusion PDEs which generate general-purpose denoising methods encompasses the homogenous heat equation, the Perona-Malik nonlinear edge-preserving diffusion equation [12], with its mathematically well-posed modification [4], and the anisotropic diffusion equation [16]. Many variants of these denoising PDEs, tailored for specific applications, have also been proposed (e.g. [18]). In the experiments reported in Sec. 4 we have used the nonlinear diffusion filter of [4]:

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

with initial condition $u(x, 0) = y(x)$ and reflecting boundary conditions, but our algorithms are applicable in the more general setting. In Eq. (1) the diffusivity $g : [0, +\infty) \rightarrow R^+$ is a decreasing function (with $g(0) = 1$, $g(r) \rightarrow 0$ while $r \rightarrow +\infty$) which favors intraregion over interregion smoothing and u_σ denotes linear convolution of u with a Gaussian kernel of standard deviation σ .

A central issue with diffusion-based denoising techniques is how to optimally select the scale at which to stop diffusing the image. At the one extreme, at a very small scale the noise has not yet been suppressed sufficiently and the estimate demonstrates 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 diffusion processes per se do not exhibit a natural stopping point. For example, the steady-state solution of nonlinear diffusion is typically the trivial constant-valued image. Even if the diffusivity function g in Eq. (1) is selected in a way that the steady-state $u(x, \infty)$ is non-trivial [2, 6], there is still no guarantee that $u(x, \infty)$ is optimal for the denoising prob-

lem in any reasonable 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 over-smoothed. The need for such explicit scale selection mechanisms has also been stressed out by other researchers in similar contexts (see e.g. the work of Lindeberg in [9]).

In our work, we emphasize that the problem of choosing the optimal scale-space snapshot can be treated 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 and prove to be particularly effective and robust in practice. Our approach is on the same spirit with [11], where similar cross-validation techniques are utilized for selecting the optimal threshold for denoising by wavelet thresholding.

The paper is organized as follows. In Sec. 2 we first formulate the problem of scale selection for denoising by non-linear diffusion and then quickly review previous work in the area. Then, in Sec. 3 we present the cross-validation model selection framework to address the problem and derive from it two practical algorithms. In Sec. 4 we describe the experiments we conducted in order to assess the performance of the proposed algorithms and compare them with other methods. Finally, in Sec. 5 we present our conclusions and give possible directions of future research.

2. Problem Formulation and Previous Work

Available is a “noisy” grayscale image y consisting of $M = M_x M_y$ pixels. We consider a lexicographic ordering of the pixels and denote by 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 , which follows the model:

$$Y_i = f^*(x_i) + N_i, \quad i = 1, \dots, M, \quad (2)$$

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 with the noisy image $y(x)$ as initial condition $u(x, 0)$, the diffusion PDE (1) generates a sequence of scale-space snapshots $f_t(x) \equiv u(x, t)$ (indexed by the scale parameter t), which are candidate models 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:

tion error made by model f_t as:

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

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 $\text{PE}(t)$ attains its minimum, i.e. $t^* = \operatorname{argmin}_{t \geq 0} \text{PE}(t)$.

We also define the *model error*:

$$\text{ME}(t) \equiv \text{ME}(f_t) = L(f^*, f_t) \quad (4)$$

In the case that the noise power tends to zero, $\text{ME}(t)$ and $\text{PE}(t)$ coincide. Otherwise the added uncertainty due to noise leads to $\text{PE}(t) > \text{ME}(t)$. For example, for square loss and zero-mean i.i.d. errors of variance σ^2 , one can easily show that $\text{PE}(t) = \text{ME}(t) + \sigma^2$. The utility of $\text{ME}(t)$ is limited in practice, since f^* is usually unknown. However one can use $\text{ME}(t)$ and the scale $t_{\text{ME}}^* = \operatorname{argmin}_{t \geq 0} \text{ME}(t)$ 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^* .

A quantity closely related to $\text{PE}(t)$ is the *extra-sample prediction error*, defined as:

$$\text{PE}_+(t) \equiv \text{PE}_+(f_t) = E_{N,X}\{L(y^{new}, f_t)\} \quad (5)$$

$\text{PE}_+(t)$ differs from $\text{PE}(t)$ because it treats the position X of pixels in the newly generated image y^{new} as random variables [5, 3]. This means that for $\text{PE}_+(t)$ 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 $\text{PE}_+(t) \geq \text{PE}(t)$. In image processing terminology, $\text{PE}_+(t)$ measures not only the uncertainty due to noise and model error, as $\text{PE}(t)$ does, but an interpolation error, as well. However, in our application we are not interested in $\text{PE}(t)$ per se but in the scale t^* that it attains its minimum. Assuming that the interpolation error terms in $\text{PE}_+(t)$ for the different models $\{f_t : t \geq 0\}$ effectively cancel out, it is plausible to take $t^* \approx t_+^*$, where $t_+^* = \operatorname{argmin}_{t \geq 0} \text{PE}_+(t)$ is the scale where $\text{PE}_+(t)$ is minimized. The gain from this assumption is that $\text{PE}_+(t)$ can be estimated directly from the noisy image y using cross-validation techniques, as we will see in Sec. 3.

A number of approaches have been proposed in the image processing community to address the problem. In [17] Weickert selects the scale t_{snr}^* which satisfies the relation $\frac{\text{var}(f_{t_{\text{snr}}^*})}{\text{var}(f_0)} = \frac{1}{1+1/\text{snr}}$, assuming that the signal-to-noise ratio snr is known and that the diffusion filter is so effective, that $f_{t_{\text{snr}}^*}$ is a good approximation of f^* .

Mrazek in [10] proposes a decorrelation criterion, selecting the scale t_{dec}^* which minimizes the correlation coefficient $t_{\text{dec}}^* = \operatorname{argmin}_{t \geq 0} \frac{\operatorname{cov}(f_0 - f_t, f_t)}{\sqrt{\operatorname{var}(f_0 - f_t)\operatorname{var}(f_t)}}$, considering $f_0 - f_t$ as “noise” and $f_{t_{\text{dec}}^*} \approx f^*$. However the decorrelation criterion does not seem to be connected with any other criterion of filtering quality, which was also noted in [10]. Finally, Solo in [13], assuming i.i.d. gaussian noise of known variance and quadratic penalty, derives a SURE-based [14] criterion for selecting the scale that minimizes $\operatorname{PE}(t)$ 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 $\operatorname{PE}_+(t)$ of Eq. (5) in a non-parametric, data-driven way [15]. Cross-validators 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 quadratic loss and linear model settings [7]. Therefore cross-validation is particularly appropriate for scale selection in our context of denoising by nonlinear diffusion 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 [1] (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. [11]).

In order to estimate $\operatorname{PE}_+(t)$ by cross-validation, we need to properly resample the noisy data $\mathcal{D} = \{(x_i, y_i) : 1 \leq i \leq M\}$ 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 $\widehat{\operatorname{PE}}_+(f_t^{-\mathcal{T}_1}) = L(y^{\mathcal{T}_1}, f_t^{-\mathcal{T}_1})$, since \mathcal{T}_1 is *independent* from the data used to build the model. We repeat this K times for different subsets $\mathcal{T}_1, \dots, \mathcal{T}_K$ of roughly the same size and average the results to get:

$$\widehat{\operatorname{PE}}_+(t) = \operatorname{PE}^{CV}(t) = \frac{1}{K} \sum_{k=1}^K \widehat{\operatorname{PE}}_+(f_t^{-\mathcal{T}_k}) \quad (6)$$

In order to derive practical algorithms for our application, we need to specify the $\{\mathcal{T}_i : 1 \leq i \leq K\}$. This issue has attracted considerable attention in the literature (see

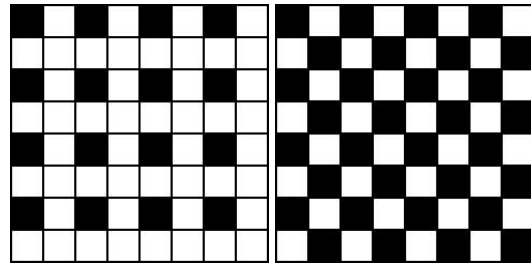


Figure 1. Resampling configurations. \mathcal{T}_1 is depicted in white. *Left:* Quadruple-CV. *Right:* Double-CV.

e.g. [8] and the references therein). 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 [7]. However, adopting *leave-one-out* cross-validation in our case would require building nonlinear diffusion scale-spaces of M images of size $(M-1)$ -pixels each, which clearly is unacceptably expensive. Hence we have explored two other data resampling configurations for our problem. Fig. 1 depicts for each of these two alternative strategies 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 $i = 1$. The first of these configurations (called *quadruple-cv* from now on), which was also used by Nason in the context of wavelet shrinkage [11], creates $K = 4$ nonlinear diffusion scale-spaces, each built up using $\approx 1/4$ of the noisy image data and used to estimate the prediction error of the model on the remaining $|\mathcal{T}_i| \approx 3/4$ of the data. The second (called *double-cv* in the following) selects the members of \mathcal{T}_i in a chessboard-like fashion, utilizing half for building the scale-space and half for testing it, repeating $K = 2$ times.

In more detail, in the case of *quadruple-cv* configuration, 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 $\{(2i + s_i, 2j + s_j) : 0 \leq i < M_x/2, 0 \leq j < M_y/2\}$. We then build the nonlinear diffusion scale-space of each of these subsampled images, appropriately scaling the diffusion PDE. For example, if Eq.(1) is utilized, we must enforce $\tau' = \tau/4$ (for the time-step), $g'(\cdot) = g(\cdot/2)$ and $\sigma' = \sigma/2$. We subsequently use Eq. (6) to get $\operatorname{PE}_{\text{quadruple-cv}}^{CV}(t)$, computing each of the four terms $\{\widehat{\operatorname{PE}}_+(f_t^{-\mathcal{T}_k}) : 1 \leq k \leq 4\}$ as follows: We take the $t' = t/4$ snapshot of the corresponding auxiliary scale-space, with dimensions $M_x/2$ by $M_y/2$ and

interpolate from it the values at the remaining $3M/4$ pixels $p_i \in \mathcal{T}_k$. A nonlinear, edge-preserving procedure should ideally be utilized for the interpolation, although, in practice, we have noticed little difference when using simple bilinear interpolation. We then penalize (by means of L) the difference 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 $\widehat{\text{PE}}_+(f_t^{-\mathcal{T}_k})$. The procedure just described adds automatic scale selection to nonlinear diffusion-based denoising procedures at roughly double the computational cost of plain nonlinear diffusion, since 4 auxiliary scale-spaces of size $M/4$ pixels each need to be built in parallel with the standard scale-space.

The concept in the case of *double-cv* resampling is similar and won't be described in detail. The main difference is that the pixels that build each of the auxiliary scale-spaces are not located on a rectangular lattice any more (see Fig. 1, right). Therefore, it is convenient to first 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. Since these auxiliary scale-spaces are full-sized, the PDE (6) doesn't need any rescaling. The overall cost of the procedure is three times the cost of the standard scale-space, since two auxiliary full-sized scale-spaces evolve in parallel with the main one.

4. Experiments and comparisons

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

In order to systematically assess the performance of the proposed algorithms and compare them with existing techniques, we 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 have proposed (*double-cv* and *quadruple-cv*), we have also implemented the *snr*-based method of We-

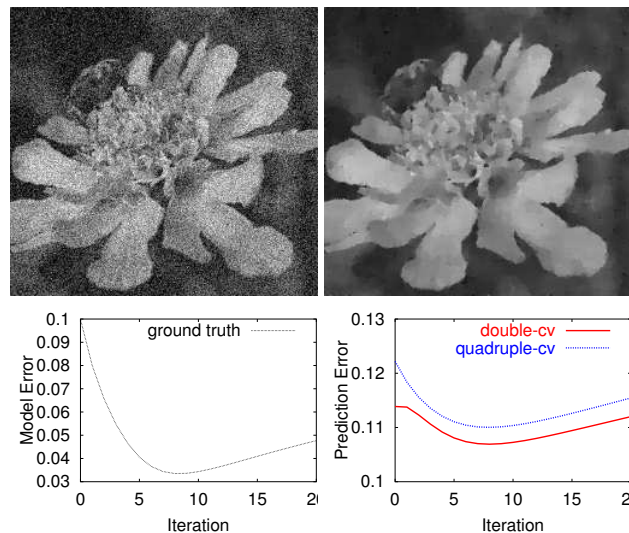


Figure 2. Image denoising example. *Up:* The noisy image and its denoised version at an automatically selected scale. *Down:* Model Error $\text{ME}(t)$ (ground truth) and Prediction Error $\text{PE}_+(t)$ (estimated by each of the two cross-validation algorithms).

ickert (with ground-truth *snr* value), and the decorrelation method of Mrazek, in the sequel dubbed *snr* and *dec*, respectively. The nonlinear diffusion scale-space used in the experiments reported here was generated by the PDE of Eq. (1) with $\sigma = 0.1$ pixels. We employed the diffusivity function $g(r) = 1/[1 + (r/\lambda)^2]$ [12], with $\lambda = 0.01$ in all of the tests (the intensity values of the images were in $[0, 1]$) and the AOS numerical scheme [16]. Bilinear interpolation was utilized for upsampling.

We experimented with three different noise types $nt \in \{\text{gaussian, salt\&pepper, speckle}\}$. 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 evaluate 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 performance results of each algorithm on the different images. We totally performed $36 \cdot 39 = 1404$ experiments. At each experiment and for each scale selection algorithm $alg \in \{\text{double-cv, quadruple-cv, dec, snr}\}$, we evaluated two quantities, namely the *relative increase in model*

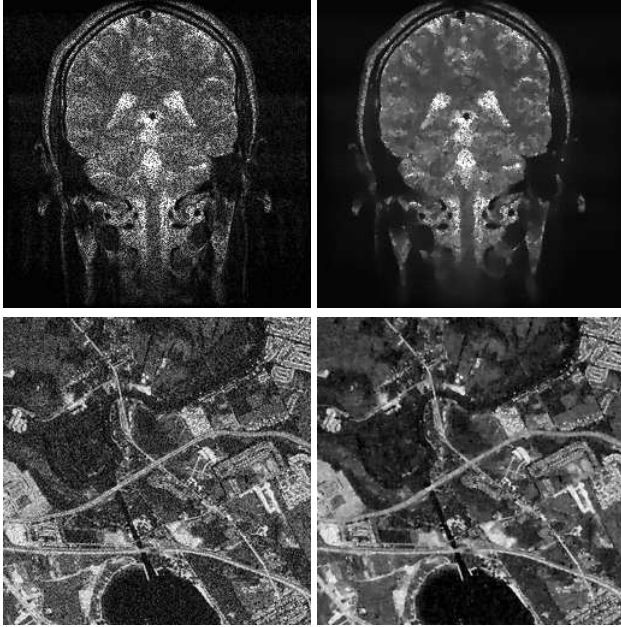


Figure 3. Examples of noisy images and their denoised versions at automatically selected scales. *Up:* An MRI scan degraded by speckle noise. *Down:* An aerial image with added gaussian noise.

error, defined as $\frac{\Delta(\text{ME})}{\text{ME}} = \frac{\text{ME}(t_{\text{alg}}^*) - \text{ME}(t_{\text{ME}}^*)}{\text{ME}(t_{\text{ME}}^*)}$ and the relative scale error, given by $\frac{\Delta t^*}{t^*} = \frac{|t_{\text{alg}}^* - t_{\text{ME}}^*|}{t_{\text{ME}}^*}$.

In Figs. 4 and 5 we present average benchmark scores for $\frac{\Delta(\text{ME})}{\text{ME}}$ and $\frac{\Delta t^*}{t^*}$, respectively, acquired by the experimental process just described. In these figures one can see that the two algorithms based on cross-validation consistently outperform the *dec* and *snr* methods. The robustness of both cross-validation algorithms, irrespectively of the noise type or the utilized loss function is particularly noteworthy.

In more detail, as far as $\frac{\Delta(\text{ME})}{\text{ME}}$ is concerned (Fig. 4), *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 $\frac{\Delta(\text{ME})}{\text{ME}}$ of the two cross-validation algorithms is almost always less than 1%, approaching sometimes 0.1%, practically meaning that the automatically selected scale t_{CV}^* in these cases coincided with the ground truth optimal scale for almost every image in the dataset. The performance of *double-cv* and *quadruple-cv* in the two cases of non-gaussian noise is less impressive, although the error still remains under 10%. Among the other two algorithms, *snr* seems to perform better than *dec*. It is noteworthy in Fig. 4 that the error made by the *snr* algorithm consistently lies around 10%. The decorrelation algo-

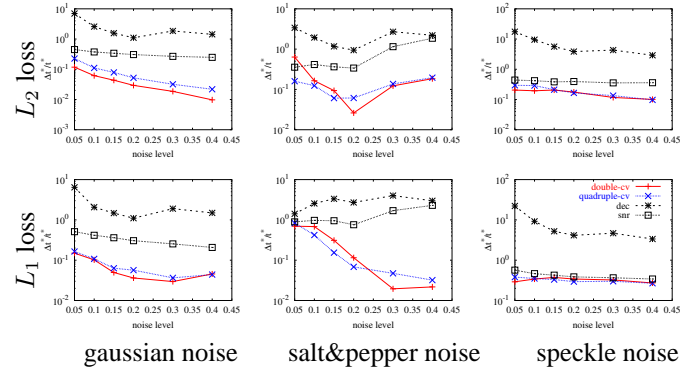


Figure 5. Relative scale error $\Delta t^*/t^*$ at various noise levels nl for the scale selection algorithms under comparison. Average results over 39 images. Log-scale on y-axis. Legend on the bottom-right figure.

gorithm *dec* performed rather poorly in our experiments. Note that the experiments in [10] also imply that *dec* doesn't perform well with the nonlinear diffusion PDE (1). Similar remarks can be made about the $\frac{\Delta t^*}{t^*}$ results shown in Fig. 5.

Finally, regarding the relative performance of the two cross-validation algorithms, *double-cv* seems from the results shown in the figures to perform slightly better than *quadruple-cv*, but the margin is too small to allow for safe conclusions, especially if we take into account that *quadruple-cv* is computationally cheaper.

5. Conclusions

This paper has presented a promising approach to automatic scale selection in nonlinear diffusion scale-spaces for denoising purposes. The problem has been formulated in a statistical model selection framework and cross-validation techniques have been utilized to estimate the optimal stopping time in a principled way. Two practical algorithms have been derived under this approach, which can be used with whatever loss function is most applicable to a specific application under consideration. Extensive experimental results on artificially degraded images have demonstrated the efficacy of the method in a range of noise types and loss functions. It will be interesting to test the applicability of the method in denoising methodologies which utilize broader classes of diffusion or morphological scale spaces.

6. Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments. Our work has been partially funded by the European Network of Excellence MUSCLE

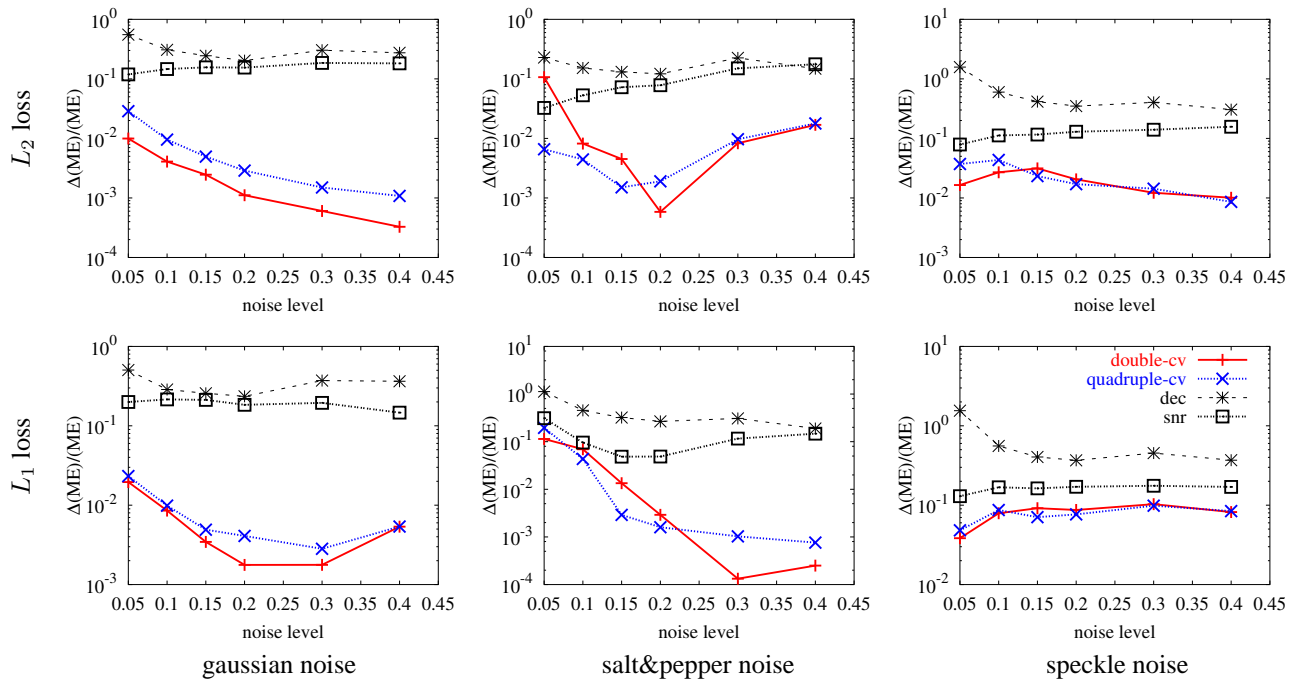


Figure 4. Relative increase in model error $\Delta(\text{ME})/\text{ME}$ as a function of the noise level nl for the scale selection algorithms under consideration. Each graph corresponds to different noise type–loss function pair. Average results computed over 39 images. Log-scale on y-axis. Legend on the bottom-right figure.

and by the Greek research programs ‘Pythagoras’ and ‘Protagoras’. GP has also been supported by the Greek State Scholarships Foundation (IKY).

References

- [1] N. S. Altman. Kernel smoothing of data with correlated errors. *J. Am. Stat. Assoc.*, 85(411):749–759, 1990.
- [2] M. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. on Image Proc.*, 7(3):421–432, 1998.
- [3] L. Breiman. The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. *J. Am. Stat. Assoc.*, 87(419):738–754, 1992.
- [4] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal Numer. Anal.*, 29(1):182–193, 1992.
- [5] B. Efron. How biased is the apparent error rate of a prediction rule? *J. Am. Stat. Assoc.*, 81(394):461–470, 1986.
- [6] 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 Image Proc.*, volume III, pages 134–137, 2001.
- [7] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [8] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proc. Int’l Joint Conf. on Artif. Intel.*, pages 1137–1143, 1995.
- [9] T. Lindeberg. Feature detection with automatic scale selection. *Int’l J. of Comp. Vision*, 30(2):79–116, 1998.
- [10] P. Mrazek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *Int’l J. of Comp. Vision*, 52(2/3):189–203, 2003.
- [11] G. P. Nason. Wavelet shrinkage using cross-validation. *J. R. Statist. Soc. B*, 58(2):463–479, 1996.
- [12] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [13] V. Solo. Automatic stopping criterion for anisotropic diffusion. In *Proc. Int’l Conf. Ac., Speech and Signal Proc.*, volume 6, pages 3929–3932, 2001.
- [14] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Stat.*, 9(6):1135–1151, 1981.
- [15] M. Stone. Cross-validated choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc. B*, 36:111–147, 1974.
- [16] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, 1998.
- [17] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201–212, 1999.
- [18] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Trans. on Image Proc.*, 11(11):1260–1270, 2002.