In M. Kerckhove (ed.), *Scale-Space and Morphology in Computer Vision. Third International Conference, Scale-Space 2001*, Vancouver, Canada, pp. 290-298.

Selection of Optimal Stopping Time for Nonlinear Diffusion Filtering

Pavel Mrázek*

Center for Machine Perception
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Praha 6, Czech Republic
www: http://cmp.felk.cvut.cz
e-mail: mrazekp@cmp.felk.cvut.cz

Abstract. We develop a novel time-selection strategy for iterative image restoration techniques: the stopping time is chosen so that the correlation of signal and noise in the filtered image is minimised. The new method is applicable to any images where the noise to be removed is uncorrelated with the signal; no other knowledge (e.g. the noise variance, training data etc.) is needed. We test the performance of our time estimation procedure experimentally, and demonstrate that it yields near-optimal results for a wide range of noise levels and for various filtering methods.

1 Introduction

If we want to restore noisy images using some method which starts from the input data and creates a set of possible filtered solutions by gradually removing noise and details from the data, the crucial question is when to stop the filtering in order to obtain the optimal restoration result. The restoration procedures needing such a decision include the linear scale space [3], the nonlinear diffusion filtering [6,1], and many others. We employ a modified version of the Weickert's edge-enhancing anisotropic diffusion [9] for most experiments in this paper.

The stopping time T has a strong effect on the diffusion result. Its choice has to balance two contradictory motivations: small T gives more trust to the input data (and leaves more details and noise in the data unfiltered), while large T means that the result becomes dominated by the (piecewise) constant model which is inherent in the diffusion equations. The scale-space people often set T to a large value (ideally infinity) and observe how the diffused function evolves with time (and converges to a constant value). As we are more concerned with image restoration and we want to obtain nontrivial results from the diffusion filter, we will have to pick a single (finite) time instant T and stop the diffusion evolution there.

We work with the following model (see Fig. 1): let $\widetilde{\mathbf{f}}$ be an ideal, noise-free (discrete) image; this image is observed by some imprecise measurement device

* This research was supported by the Czech Ministry of Education under project LN00B096.

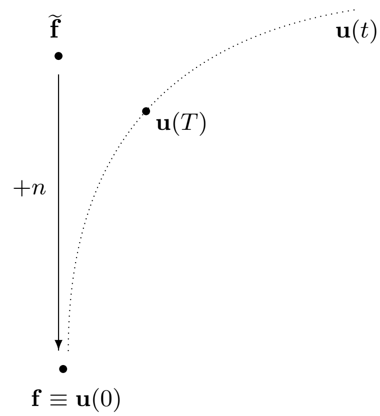


Fig. 1. Model of the time-selection problem for the diffusion filtering. We want to select the filtered image $\mathbf{u}(T)$ which is as close as possible to the ideal signal $\widetilde{\mathbf{f}}$.

to obtain an image $\mathbf{f}$. We assume that some noise n is added to the signal during the observation so that

$$\mathbf{f} = \widetilde{\mathbf{f}} + n. \tag{1}$$

Furthermore, we assume that the noise n is uncorrelated with the signal $\widetilde{\mathbf{f}}$, and that the noise has zero mean value, $E(n) = 0$.[1]

The diffusion filtering starts with the noisy image as its initial condition, $\mathbf{u}(0) = \mathbf{f}$, and the diffusion evolves along some trajectory $\mathbf{u}(t)$. This trajectory depends on the diffusion parameters and on the input image; the optimistic assumption is that the noise will be removed from the data before any important features of the signal commence to deteriorate significantly, so that the diffusion leads us somewhere 'close' to the ideal data. This should be the case if the signal adheres to the piecewise constant model inherent in the diffusion equation.

The task of the stopping time selection can be formulated as follows: select that point $\mathbf{u}(T)$ of the diffusion evolution which is nearest to the ideal signal $\widetilde{\mathbf{f}}$. Obviously, the ideal signal is normally not available; the optimal stopping time T can only be estimated by some criteria, and the distance[2] between the ideal and the filtered data serves only in the experiments to evaluate the performance of the estimation procedure.

[1] Let us review the statistical definitions used in the paper (see e.g. Papoulis [5]). For the statistical computations on images, we treat the pixels of an image as independent observations of a random variable.
The *mean* or *expectation* of a vector x is $\overline{x} = E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$.
We define the *variance* of a signal x as $\text{var}(x) = E\left[(x - \overline{x})^2\right]$.
The *covariance* of two vectors x, y is given by $\text{cov}(x, y) = E\left[(x - \overline{x}) \cdot (y - \overline{y})\right]$.
The normalized form of the covariance is called the *correlation coefficient*, $\text{corr}(x, y) = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}}$.

[2] In the experiments below, we measure the distance of two images by the *mean absolute deviation*, $\text{MAD}(x - y) = E(|x - y|)$.

In the following paragraphs we first cite the approaches to stopping time selection which have appeared in the literature, and comment on them. Then we develop a novel and reliable time-selection strategy based on signal–noise decorrelation.

2 Previous work

In the diffusion model of Catté *et al.* [1], the image gradient for the diffusivity computation is regularized by convolution with a Gaussian smoothing kernel G_σ. The authors argue that this regularization introduces a sort of time: the result of convolution is the same as the solution to the linear heat equation at time $t = \frac{\sigma^2}{2}$, so it is coherent to correlate the stopping time T and the 'time' of the linear diffusion. However, the equality $t = \frac{\sigma^2}{2}$ is rather a lower estimate of the stopping time: because of the diffusion process inhibited near edges, the nonlinear diffusion is always slower than the linear one, and needs a longer time to reach the desired results.

Dolcetta and Ferretti [2] recently formulated the time selection problem as a minimization of the functional

$$E(T) = \int_0^T E_c + E_s \tag{2}$$

where E_c is the computing cost and E_s the stopping cost, the latter encouraging filtering for small T. The authors provide a basic example $E_c = c$, $E_s = -\left(\int_\Omega |\mathbf{u}(x,T) - \mathbf{u}(x,0)|^2 dx\right)^2$ where the constant c balancing the influence of the two types of costs has to be computed from a typical image to be filtered.

Sporring and Weickert in [7] study the behaviour of generalized entropies, and suggest that the intervals of minimal entropy change indicate especially stable scales with respect to evolution time. They estimate that such scales could be good candidates for stopping times in nonlinear diffusion scale spaces. However, as the entropy can be stable on whole *intervals*, it may be difficult to decide on a single stopping instant from that interval; we are unaware of their idea being brought to practice in the field of image restoration.

Weickert mentioned more ideas on the stopping time selection, more closely linked to the noise-filtering problem, in [10]. They are based on the notion of relative variance.

The variance $\mathrm{var}(\mathbf{u}(t))$ of an image $\mathbf{u}(t)$ is monotonically decreasing with t and converges to zero as $t \to \infty$. The *relative variance*

$$r(\mathbf{u}(t)) = \frac{\mathrm{var}(\mathbf{u}(t))}{\mathrm{var}(\mathbf{u}(0))} \tag{3}$$

decreases monotonically from 1 to 0 and can be used to measure the distance of $\mathbf{u}(t)$ from the initial state $\mathbf{u}(0)$ and the final state $\mathbf{u}(\infty)$. Prescribing a certain value for $r(\mathbf{u}(T))$ can therefore serve as a criterion for selection of the stopping time T.

Let again $\widetilde{\mathbf{f}}$ be the ideal data, the measured noisy image $\mathbf{f} = \widetilde{\mathbf{f}} + n$, and let the noise n be of zero mean and uncorrelated with $\widetilde{\mathbf{f}}$. Now assume that we know the variance of the noise, or (equivalently, on the condition that the noise and the signal are uncorrelated) the *signal-to-noise ratio*, defined as the ratio between the original image variance and the noise variance,

$$\mathrm{SNR} \equiv \frac{\operatorname{var}\left(\widetilde{\mathbf{f}}\right)}{\operatorname{var}(n)}. \tag{4}$$

As the signal $\widetilde{\mathbf{f}}$ and the noise n are uncorrelated, we have

$$\operatorname{var}(\mathbf{f}) = \operatorname{var}\left(\widetilde{\mathbf{f}}\right) + \operatorname{var}(n). \tag{5}$$

Substituting from this equality for $\operatorname{var}(n)$ into (4), we obtain by simple rearrangement that

$$\frac{\operatorname{var}\left(\widetilde{\mathbf{f}}\right)}{\operatorname{var}(\mathbf{f})} = \frac{1}{1 + \frac{1}{\mathrm{SNR}}}. \tag{6}$$

We take the noisy image for the initial condition of our diffusion filter, $\mathbf{u}(0) = \mathbf{f}$. An ideal diffusion filter would first eliminate the noise before significantly affecting the signal; if we stop at the right moment, we might substitute the filtered data $\mathbf{u}(T)$ for the ideal signal $\widetilde{\mathbf{f}}$ in (6). Relying on this analogy, we can choose the stopping time T such that the relative variance satisfies

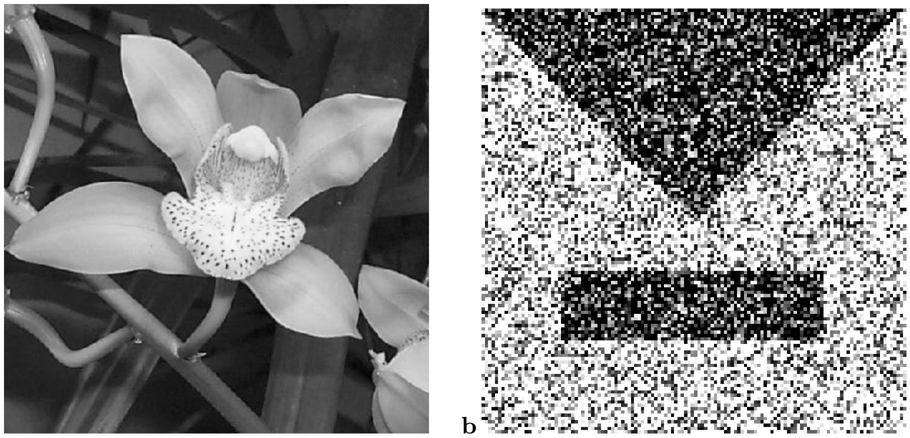
$$r(\mathbf{u}(T)) = \frac{\operatorname{var}(\mathbf{u}(T))}{\operatorname{var}(\mathbf{u}(0))} = \frac{1}{1 + \frac{1}{\mathrm{SNR}}} \tag{7}$$

Weickert remarks that the criterion (7) tends to underestimate the optimal stopping time, as even a well-tuned filter cannot avoid influencing the signal before eliminating the noise.

So far the Weickert's suggestions from [10]: knowing the SNR, we decide to filter the image until some distance from the noisy data is reached, and the formula (7) tells us when to stop the diffusion. This idea seems natural and resembles also that used in the total variation minimizing methods (see overview in [9, pp. 50-52]). However, our experiments indicate that this approach does not usually yield the optimal stopping time. Let us study in more detail why the problems occur.

3 Decorrelation criterion

The equality (5) and hence the equation (6) are valid only if the signal and the noise are uncorrelated. This assumption holds for $\widetilde{\mathbf{f}}$ and n, but not necessarily for the filtered signal $\mathbf{u}(T)$ and the difference $\mathbf{u}(0) - \mathbf{u}(T)$; the latter is needed for the equation (7) to be justified. In other words (if we substitute mentally the filtered function $\mathbf{u}(T)$ for $\widetilde{\mathbf{f}}$, the difference $n_u \equiv \mathbf{u}(0) - \mathbf{u}(T)$ for the noise n, and $\mathbf{u}(0)$ for $\mathbf{f}$ in (5) and (6)), the formula (7) is useful only if the random variables $\mathbf{u}(T)$ and $(\mathbf{u}(0) - \mathbf{u}(T))$ are uncorrelated.



a **b**

Fig. 2. Experimental data. Left: Cymbidium image (courtesy Michal Haindl). Right: Noisy input image ($[0,127]^2 \to [0,255]$) for the 'Triangle and rectangle' experiment. Noise with uniform distribution in the range $[-255,255]$ was added to two-valued synthetic data.

Inspired by these observations, we arrive to the following idea: if the unknown noise n is uncorrelated with the unknown signal $\widetilde{\mathbf{f}}$, wouldn't it be reasonable to minimize the covariance of the 'noise' $(\mathbf{u}(0) - \mathbf{u}(t))$ with the 'signal' $\mathbf{u}(t)$, or – better – employ its normalized form, the correlation coefficient

$$\operatorname{corr}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big) = \frac{\operatorname{cov}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big)}{\sqrt{\operatorname{var}\big(\mathbf{u}(0) - \mathbf{u}(t)\big) \cdot \operatorname{var}\big(\mathbf{u}(t)\big)}} \tag{8}$$

and choose the stopping time T so that the expression (8) is as small as possible? This way, instead of determining the stopping time so that $(\mathbf{u}(0) - \mathbf{u}(T))$ satisfies a quantitative property and its variance is equal to the known variance of the noise n, we try to enforce a qualitative feature: if the ideal $\widetilde{\mathbf{f}}$ and n were uncorrelated, we require that their artificial substitutes $\mathbf{u}(T)$ and $(\mathbf{u}(0) - \mathbf{u}(T))$ reveal the same property, to the extent possible, and select

$$T = \arg\min_t \operatorname{corr}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big). \tag{9}$$

Let us test and validate this new stopping time criterion experimentally.

4 Experiments

We added various levels of Gaussian noise to the cymbidium image shown in Fig. 2 left, filtered by nonlinear diffusion (more precisely a modified version of the Weickert's edge-enhancing anisotropic diffusion [9], numerically implemented

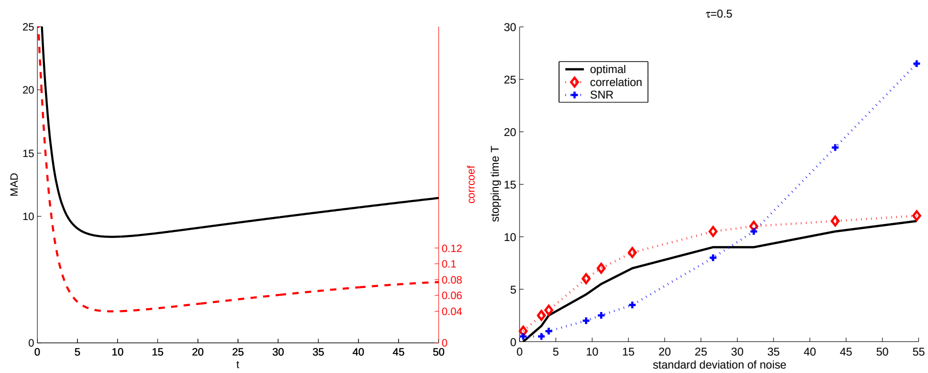


Fig. 3. Left: The distance $\mathrm{MAD}(\mathbf{u}(t) - \widetilde{\mathbf{f}})$ (solid line) and the correlation coefficient $\mathrm{corr}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big)$ (dashed line) developing with the diffusion time. Right: The stopping time T_{SNR} determined by the SNR method (dotted with crosses), and T_{corr} obtained through the covariance minimization (dotted with diamonds) compared to the optimal stopping time T_{opt} (solid line). The graphs are plotted against the standard deviation of noise in the input image.

using the AOS scheme [8]), and observed how the signal–noise correlation measured by equation (8) develops with the diffusion time. A typical example is drawn in Fig. 3 left: you can observe that the plot of the MAD criterion of the filtering quality coincides very well with the graph of the correlation coefficient $\mathrm{corr}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big)$.

A more thorough study of the performance of the stopping time selection criteria (measured again on the cymbidium data) is seen in figures 3 right and 4. The former compares three stopping times: the optimal T_{opt} is the time instant for which the filtered image $\mathbf{u}(t)$ is closest to the noise-free $\widetilde{\mathbf{f}}$ in the MAD distance; obviously, T_{opt} can be found only in the artificial experimental setting, the noise-free $\widetilde{\mathbf{f}}$ is normally not available. The second stopping time T_{SNR} is determined using the criterion (7) (which requires the knowledge of the noise variance or SNR). The stopping time T_{corr} minimizes the correlation coefficient of equation (8). All alternative stopping times are computed for a series of input images with varied amount of noise present. While the SNR method easily underestimates or overestimates the optimal stopping time (depending on the amount of noise in the input data), the correlation minimization leads to near-optimal results for all noise levels. The graph is plotted for iteration time step $\tau = 0.5$, other choices $\tau \in \{0.1,\ 1\}$ gave similar results.

The actually obtained quality measure $\mathrm{MAD}(\mathbf{u}(T) - \widetilde{\mathbf{f}})$ is shown in Fig. 4, again with $\tau = 0.5$. You can see that for all noise levels the correlation-estimated time leads to filtering results very close to the optimal values obtainable by the nonlinear diffusion.

Let us return for a moment to Fig. 3 left. At the beginning of the diffusion filtering, the correlation coefficient declines fast until it reaches its minimum.

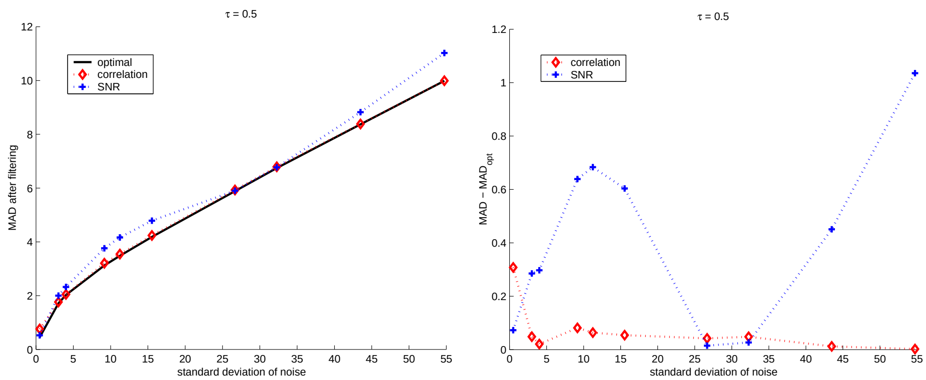


Fig. 4. Left: the MAD distance of the filtered data from the ideal noise-free image, $\mathrm{MAD}(\mathbf{u}(T) - \widetilde{\mathbf{f}})$, using the SNR and the correlation-minimization time selection strategies. Right: the difference between the estimated result and the optimal one, $\mathrm{MAD}(\mathbf{u}(T) - \mathbf{u}(T_{\mathrm{opt}}))$.

If for some data the graph behaves differently, it may serve as a hint on some problems. As an example, we observed that if there is only a small amount of noise in the input image, the correlation $\mathrm{corr}\big(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\big)$ might grow from the first iterations. In such a case, the iteration time step τ has to be decreased adaptively and the diffusion restarted from time $t = 0$ until the correlation plot exhibits a clear minimum.

Another experiment compares the results of different diffusion algorithms filtering an originally black and white image with non-Gaussian additive noise. The input data are shown in Fig. 2 right: the noisy image was obtained by adding noise of uniform distribution in the range $[-255, 255]$ to the ideal input, and by restricting the noisy values into the interval $[0, 255]$.

In Fig. 5, the noise is smoothed by linear diffusion, isotropic nonlinear diffusion, and two anisotropic diffusion filters; the grey-values are stretched to the whole interval $[0, 255]$ so that a higher contrast between the dark and bright regions corresponds to a better noise-filtering performance. In all cases, the stopping time was determined autonomously by the signal–noise decorrelation criterion (9). You can see that in all cases, although quite different filtering algorithms were employed, the stopping criterion leads to results where most of the noise is removed and the ideal signal becomes apparent or suitable for further processing; we support this statement by showing the thresholded content of the filtered images in Fig. 6.

The stopping criterion was designed to minimize the MAD distance from the ideal function. If visual quality was the goal to be achieved, we would probably stop the diffusion later, especially as linear diffusion (Fig. 5a) and the Weickert's edge-enhancing anisotropic diffusion [9] with maximum amount of diffusion in the coherence direction ($\varphi_2 = 1$, Fig. 5c) are concerned. We find however that the MAD distance and the visual quality are in a good agreement in Fig. 5d which

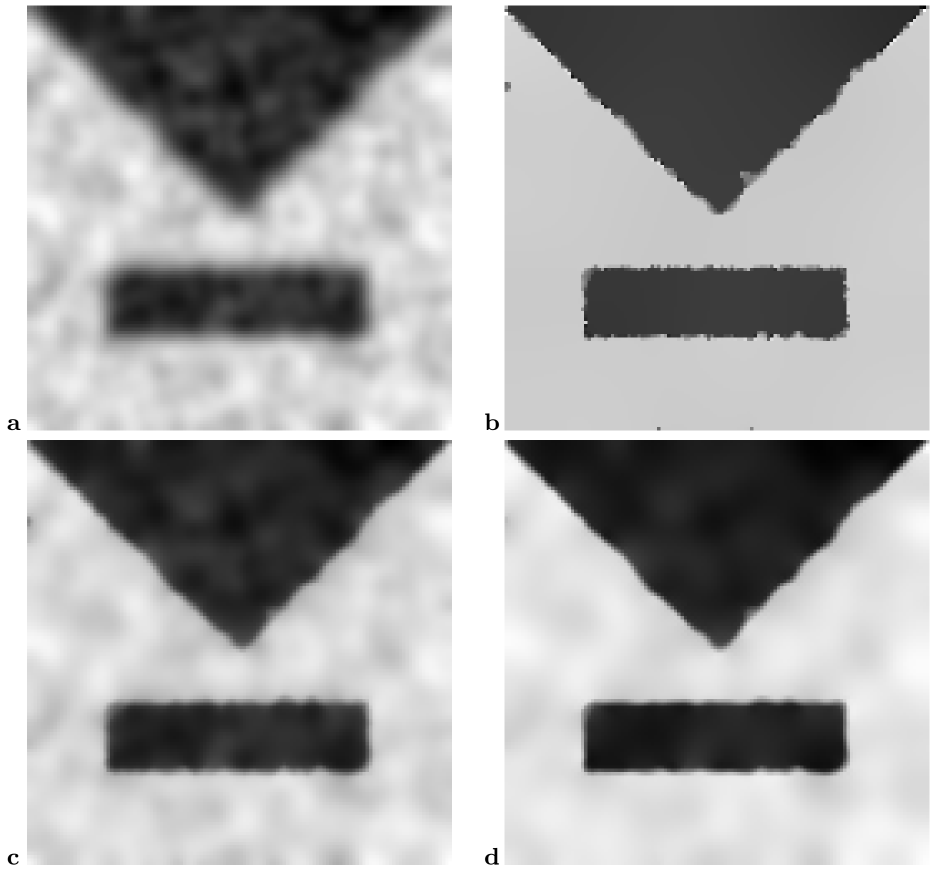


Fig. 5. Comparing the different diffusion algorithms on the noisy data of Fig. 2 right, all with the stopping time selected autonomously by minimizing the criterion (8): (a) linear diffusion, $T = 3.8$; (b) isotropic nonlinear diffusion, $T = 125$; (c) anisotropic NL diffusion, $\varphi_2 = 1$, $T = 15$; (d) anisotropic NL diffusion, $\varphi_2 = 0.2$, $T = 32$.
In (b)–(d), the parameters $\sigma = 1$, $\tau = 1$ were employed, and the parameter λ was estimated using the Perona-Malik procedure from percentile $p = 0.9$ in each step.

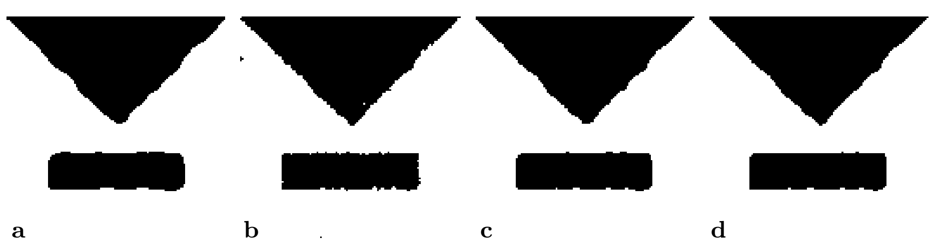


Fig. 6. Thresholded versions of the images in Fig. 5

represents the result of the edge-enhancing diffusion with a smaller amount of diffusion in the coherence direction, $\varphi_2 = 0.2$. Because of limited space, we have to refer the reader to Pavel Mrázek's thesis [4] for details on the filtering procedures and for more experimental results verifying the decorrelation criterion.

5 Conclusion

We have developed a novel method to estimate the optimal stopping time for iterative image restoration techniques such as nonlinear diffusion. The stopping time is chosen so that the correlation of signal $\mathbf{u}(T)$ and 'noise' $(\mathbf{u}(0) - \mathbf{u}(T))$ is minimised. The new criterion outperforms other time selection strategies and yields near-optimal results for a wide range of noise levels and filtering parameters. The decorrelation criterion is also more general, being based only on the assumption that the noise and the signal in the input image are uncorrelated; no knowledge on the variance of the noise, and no training images are needed to tune any parameters of the method.

Acknowledgements

The author would like to express his thanks to Mirko Navara for helpful comments on the manuscript.

References

1. F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge-detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, 1992.
2. I. Capuzzo Dolcetta and R. Ferretti. Optimal stopping time formulation of adaptive image filtering. *Applied Mathematics and Optimization*, 2000. To appear.
3. Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
4. Pavel Mrázek. *Nonlinear Diffusion for Image Filtering and Monotonicity Enhancement*. PhD thesis, Czech Technical University, 2001. To appear.
5. Athanasios Papoulis. *Probability and Statistics*. Prentice-Hall, 1990.
6. P. Perona and J. Malik. Scale-space and edge-detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
7. Jon Sporring and Joachim Weickert. Information measures in scale-spaces. *IEEE Transactions on Information Theory*, 45:1051–1058, 1999.
8. J. Weickert, B.M. ter Haar Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7:398–410, 1998.
9. Joachim Weickert. *Anisotropic Diffusion in Image Processing*. European Consortium for Mathematics in Industry. B. G. Teubner, Stuttgart, 1998.
10. Joachim Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201–212, 1999.