International Journal of Computer Vision 52(2/3), 189–203, 2003

Selection of Optimal Stopping Time for Nonlinear Diffusion Filtering

PAVEL MRÁZEK
Center for Applied Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Technická 2, 166 27 Praha 6, Czech Republic
mrazekp@cmp.felk.cvut.cz; www: http://cmp.felk.cvut.cz/~mrazekp

MIRKO NAVARA
Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Technická 2, 166 27 Praha 6, Czech Republic
navara@cmp.felk.cvut.cz; www: http://cmp.felk.cvut.cz/~navara

Received January 18, 2002; Revised August 5, 2002; Accepted November 22, 2002

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 minimized. The new method is applicable to any images where the noise to be removed is uncorrelated with the signal, under the assumptions that the filter used is suitable for the given type of data, and that neither the additive noise nor the filtering procedure alter the average gray value; no other knowledge (e.g. the noise variance, training data etc.) is needed.

We analyse the theoretical properties of the method, then 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.

Keywords: nonlinear diffusion, stopping time, regularization parameter, image restoration, scale space

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 linear scale space (Iijima, 1962; Witkin, 1983), nonlinear diffusion filtering (Perona and Malik, 1990; Catté et al., 1992), anisotropic diffusion filtering (Weickert, 1998), and many others.

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 model of the filtering procedure—piecewise constant for nonlinear diffusion, smooth functions for linear scale space, etc.

In scale-space theory, people often set the stopping time 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 $\tilde{\mathbf{f}}$ be an ideal, noise-free (discrete) image; this image is observed by some imprecise measurement device to obtain an image $\mathbf{f}$. We assume that some

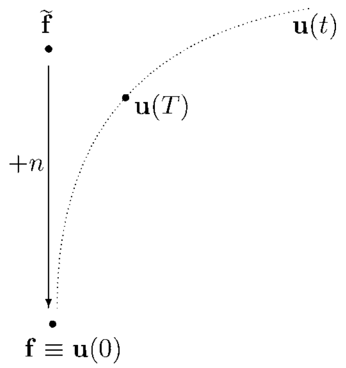

Figure 1. Model of the time-selection problem for the diffusion filtering. We want to select the filtered image $\mathbf{u}(T)$ which is as close as possible to the ideal signal $\tilde{\mathbf{f}}$.

noise n is added to the signal during the observation so that

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

Furthermore, we assume that the noise n is uncorrelated with the signal $\tilde{\mathbf{f}}$, and that the noise has zero mean value, $E(n) = 0$ (see Appendix for a review of statistical formulas and notation used in the paper).

The diffusion filtering starts with the noisy image as its initial condition, $\mathbf{u}(0) = \mathbf{f}$, and the diffusion evolves along some trajectory $\mathbf{u}(t)$, $t \in [0, \infty)$. 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 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 $\tilde{\mathbf{f}}$. Obviously, the ideal signal is normally not available; the optimal stopping time T can only be estimated by some criteria, and the distance between the ideal and the filtered data serves only in the experiments to evaluate the performance of the estimation procedure.

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

2. Previous Work

In the diffusion model of Catté et al. (1992), 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. Also, the authors do not address the question of choosing a good value for the parameter σ.

Dolcetta and Ferretti (2001) recently formulated the time selection problem as a minimization of the functional

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

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

$$E_c = c \tag{3}$$

$$E_s = -\left(\int_\Omega |\mathbf{u}(x, T) - \mathbf{u}(x, 0)|^2\, dx \right)^2 \tag{4}$$

where the constant c balancing the influence of the two types of costs has to be computed from a typical image to be filtered. Typically, the distance between filtered and original data increases faster at the early stage of the diffusion process. With the Eqs. (2)–(4), the filtering will continue until this increase in distance becomes outweighted by the constant in the computing cost E_c.

Sporring and Weickert (1999) 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 into 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 Weickert (1999). 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*

$$r(\mathbf{u}(t)) = \frac{\mathrm{var}(\mathbf{u}(t))}{\mathrm{var}(\mathbf{u}(0))} \tag{5}$$

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 $\tilde{\mathbf{f}}$ be the ideal data, the measured noisy image $\mathbf{f} = \tilde{\mathbf{f}} + n$, and let the noise n be of zero mean and uncorrelated with $\tilde{\mathbf{f}}$. 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,

$$\mathrm{SNR} \equiv \frac{\mathrm{var}(\tilde{\mathbf{f}})}{\mathrm{var}(n)}. \tag{6}$$

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

$$\mathrm{var}(\mathbf{f}) = \mathrm{var}(\tilde{\mathbf{f}}) + \mathrm{var}(n). \tag{7}$$

Substituting from this equality for $\mathrm{var}(n)$ into (6), we obtain by simple rearrangement that

$$\frac{\mathrm{var}(\tilde{\mathbf{f}})}{\mathrm{var}(\mathbf{f})} = \frac{1}{1 + \frac{1}{\mathrm{SNR}}}. \tag{8}$$

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 $\tilde{\mathbf{f}}$ in (8). Relying on this analogy, we can choose the stopping time T such that the relative variance satisfies

$$r(\mathbf{u}(T)) = \frac{\mathrm{var}(\mathbf{u}(T))}{\mathrm{var}(\mathbf{u}(0))} = \frac{1}{1 + \frac{1}{\mathrm{SNR}}} \tag{9}$$

So far the Weickert's suggestions from Weickert (1999): knowing the SNR (or, equivalently, knowing the variance of noise in the input image), we decide to filter the image until some distance from the noisy data is reached, and the formula (9) tells us when to stop the diffusion. This idea seems natural and resembles also that used in the total variation minimizing methods (Rudin et al., 1992).

Weickert remarks that the criterion (9) tends to underestimate the optimal stopping time, as even a well-tuned filter cannot avoid influencing the signal before eliminating the noise.

Remark 1. We suggest a partial explanation of this phenomenon. The points with all coordinates equal form the 'diagonal' of the space. The variance of a vector (considered as a vector of realizations of a real-valued random variable) is the square of the Euclidean distance from the diagonal. During the diffusion process, the grey level is preserved, and the whole trajectory $\mathbf{u}(t)$, $t \in [0, \infty)$, as well as $\mathbf{f}$, lie in a hyperplane, H, perpendicular to the diagonal. The point $\mathbf{u}(\infty) = \lim_{t\to\infty} \mathbf{u}(t)$ is the only point of H lying on the diagonal (because $\mathrm{var}(\mathbf{u}(\infty)) = 0$). The formula (9) determines the point on the trajectory $\mathbf{u}(t)$ which has the same distance from $\mathbf{u}(\infty)$ as $\mathbf{f}$, i.e.,

$$|\mathbf{u}(T) - \mathbf{u}(\infty)| = |\mathbf{f} - \mathbf{u}(\infty)| = \sqrt{\frac{\mathrm{var}(\mathbf{u}(0))}{1 + \frac{1}{\mathrm{SNR}}}}. \tag{10}$$

Such a point is unique because $\mathrm{var}(\mathbf{u}(t))$ is nonincreasing.

For low noise levels, criterion (9) determines a point which is more distant from $\mathbf{u}(\infty)$ than the optimum. This can be demonstrated on an extremely simplified case when the trajectory $\mathbf{u}(t)$ is the straight line segment from $\mathbf{u}(0)$ to $\mathbf{u}(\infty)$. Its point closest to $\mathbf{f}$ is the point $\mathbf{u}(T^*)$ with distance

$$|\mathbf{u}(T^*) - \mathbf{u}(\infty)| = \frac{\sqrt{\mathrm{var}(\mathbf{u}(0))}}{1 + \frac{1}{\mathrm{SNR}}},$$

i.e., strictly smaller than that given by (10). (This very special case appears when both n and $\tilde{\mathbf{f}} - \mathbf{u}(\infty)$ decrease with the same relative speed during the diffusion. Such a situation is not usual as normally the noise contains more high frequency components that are filtered first. Nevertheless, the conclusion that (9) gives a lower than optimal diffusion time remains basically valid in much more general situations.)

Our experiment confirms the above Weickert's observation only partially (see Fig. 4): criterion (9) underestimated the stopping time for low levels of noise, but overestimated the time when high level of noise was applied. Also Rem. 1 explains the imprecision of (9) only for low noise levels. Further problems are caused by correlations between signal and noise which are introduced by the filtering process. The equality (7) and hence the Eq. (8) are valid only if the signal and the noise are uncorrelated. This assumption holds for $\tilde{\mathbf{f}}$ 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 Eq. (9) to be justified. In other words (if we substitute mentally the filtered function $\mathbf{u}(T)$ for $\tilde{\mathbf{f}}$, the difference $\mathbf{u}(0) - \mathbf{u}(T)$ for the noise n, and $\mathbf{u}(0)$ for $\mathbf{f}$ in (7) and (8)), the formula (9) is useful only if $\mathbf{u}(T)$ and $\mathbf{u}(0) - \mathbf{u}(T)$ are uncorrelated.

3. Decorrelation Criterion

Let us consider once more the desired analogy between ideal and filtered data: $\tilde{\mathbf{f}}$ is the ideal signal, some noise n is added to form the observed signal $\mathbf{f} = \tilde{\mathbf{f}} + n$. The diffusion filtering starts from $\mathbf{u}(0) = \mathbf{f}$ to create a series of possible solutions $\mathbf{u}(t)$; in practice, t attains discrete values $0, t_1, t_2, \ldots$. We want to select the stopping time T so that $\mathbf{u}(T)$ is an optimal substitute for the ideal signal $\tilde{\mathbf{f}}$.

We now arrive at the main idea of this paper: if the unknown noise n is uncorrelated with the unknown signal $\tilde{\mathbf{f}}$, it could 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

$$\operatorname{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)) = \frac{\operatorname{cov}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))}{\sqrt{\operatorname{var}(\mathbf{u}(0) - \mathbf{u}(t)) \cdot \operatorname{var}(\mathbf{u}(t))}} \tag{11}$$

and choose the stopping time T so that the expression (11) is as small as possible. This way, we exploit the analogy between the ideal signal $\tilde{\mathbf{f}}$ and its filtering substitution $\mathbf{u}(T)$, but 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 (which we saw in Weickert (1999)), we try to enforce a qualitative feature: if the ideal $\tilde{\mathbf{f}}$ and n were uncorrelated, we require that their computed estimates $\mathbf{u}(T)$ and $\mathbf{u}(0) - \mathbf{u}(T)$ reveal the same property, to the extent possible, and select

$$T = \operatorname*{argmin}_t |\operatorname{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|. \tag{12}$$

We call Eq. (12) the *decorrelation criterion* for the selection of the diffusion stopping time. In the following sections we first study its theoretical properties, then test and validate this new stopping time criterion experimentally.

3.1. Theoretical Properties

The decorrelation criterion allows us to choose the diffusion stopping time without any additional knowledge and using quite mild assumptions about noise and the filtering method. As we shall show by experiments, its results are often close to optimum. Nevertheless, from the theoretical viewpoint, there are several negative results. We cannot prove that the optimum in the sense of the decorrelation criterion (12) coincides with results of some other criteria for filtering quality. Moreover, without additional assumptions on the properties of the noise, the signal and the filter, neither the filtering quality (measured by the distance $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$) nor the correlation $|\operatorname{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|$ are guaranteed to be unimodal and possess a single minimum. We will try to provide some insight into this fact by analysing the combination of the decorrelation criterion with the simplest, linear diffusion process.

As a positive result, we show that the correlation between $(\mathbf{u}(0) - \mathbf{u}(t))$ and $\mathbf{u}(t)$ remains nonnegative for a linear diffusion filter.We experienced that the decorrelation criterion provides more reliable estimates when combined with filters which keep this correlation nonnegative.

3.1.1. Combination with Linear Diffusion. Let us first address the question whether we can hope to find a unique $\mathbf{u}(T)$ which minimizes the distance to the ideal $\tilde{\mathbf{f}}$ over all $\mathbf{u}(t)$ in the diffused sequence.

The filtering procedure starts from $\mathbf{u}(0) = \tilde{\mathbf{f}} + n$. Diffusion preserves the average grey level, $E(\mathbf{u}(0))$; without loss of generality, we subtract this value from all data and assume in the sequel that $E(\mathbf{u}(0)) = 0$. In the discrete setting, the linear diffusion result at time t can be expressed in vector notation as multiplication of the initial $\mathbf{u}(0)$ by some matrix $\boldsymbol{A}(t)$,

$$\mathbf{u}(t) = \boldsymbol{A}(t)\mathbf{u}(0), \tag{13}$$

where the matrix $\boldsymbol{A}(t)$ is symmetric, has unit row sum, and its elements are nonnegative. Then, starting from the norm of $\mathbf{u}(t) - \tilde{\mathbf{f}}$ and using triangle inequality in the last step,

$$\begin{aligned}|\mathbf{u}(t) - \tilde{\mathbf{f}}| &= |\boldsymbol{A}(t)\mathbf{u}(0) - \tilde{\mathbf{f}}| = |\boldsymbol{A}(t)(\tilde{\mathbf{f}} + n) - \tilde{\mathbf{f}}| \\ &= |\boldsymbol{A}(t)\tilde{\mathbf{f}} - \tilde{\mathbf{f}} + \boldsymbol{A}(t)n| \\ &\leq |\boldsymbol{A}(t)\tilde{\mathbf{f}} - \tilde{\mathbf{f}}| + |\boldsymbol{A}(t)n|. \end{aligned} \tag{14}$$

It is known that in well-posed scale spaces the distance $|\mathbf{u}(0) - \mathbf{u}(t)|$ increases monotonically with increasing time t, and the norm of any diffused signal decreases monotonically, see Weickert (1998). Hence, Eq. (14) states that the distance $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$ is bounded by the sum of an increasing function $|\boldsymbol{A}(t)\tilde{\mathbf{f}} - \tilde{\mathbf{f}}|$ and a decreasing one, $|\boldsymbol{A}(t)n|$.

Ideally, we would like this sum to be a unimodal function of t, decreasing at the beginning and increasing later. Unfortunately, although Bertero and Boccacci (1998) claim that this is the case for linear regularization methods (linear diffusion belongs to this class), unimodality does not hold in general. Some insight into this fact can be gained if we reconsider linear diffusion in Fourier domain.

We may understand $\mathbf{u}(t)$ as a vector from $\mathbb{R}^N \subset \mathbb{C}^N$, where $N = N_1 \cdot N_2$ is the number of pixels in the $N_1 \times N_2$ image. It is uniquely determined by its image in the two-dimensional discrete Fourier transform, $\boldsymbol{U}(t) \in \mathbb{C}^N$. The discrete Fourier image is obtained by multiplication by an orthonormal $N \times N$ matrix, $\boldsymbol{H}$. This transformation preserves the inner product, therefore also the norm and angles. We may also understand $\mathbf{u}(t)$ and $\boldsymbol{U}(t)$ as two representations of the same vector in two different orthonormal bases.

The Fourier transform is linear, therefore $\boldsymbol{U}(0) = \tilde{\boldsymbol{F}} + \boldsymbol{N}$, where $\tilde{\boldsymbol{F}}, \boldsymbol{N}$ are the Fourier images of $\tilde{\mathbf{f}}$, n, respectively. Each entry of $\boldsymbol{U}(0)$ represents the amount of one frequency present in the input image, and it is again the sum of the same frequency from the ideal signal and the noise. Linear diffusion is equivalent to low-pass filtering or convolution with a Gaussian kernel. In the Fourier domain, it gradually sends the values of all the entries of $\boldsymbol{U}(t)$ to zero, attenuating higher frequencies earlier; it decreases the amplitudes and does not change the phases of the components of the Fourier series. Thus the entry $\boldsymbol{U}_{jk}(t)$ of the Fourier image $\boldsymbol{U}(t)$ with coordinates j, k is a multiple of its initial value,

$$\boldsymbol{U}_{jk}(t) = a_{jk}(t) \cdot \boldsymbol{U}_{jk}(0), \tag{15}$$

where $a_{jk}(t) \in [0, 1]$ is a real attenuation coefficient. If the bands of the input signal and noise do not overlap much (with the noise contained mostly in higher frequencies than the signal), we may be sure that linear diffusion of $\boldsymbol{U}(0)$ will lead us close to the ideal signal $\tilde{\boldsymbol{F}}$; in general, without any knowledge on the frequencies of the ideal signal and the noise, we cannot say anything about the expected filtering performance and the distance $|\boldsymbol{U}(t) - \tilde{\boldsymbol{F}}|$ does not have to be unimodal with a single minimum. These observations hold identically in the original space for $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$.

Example 1. Suppose that the input $\tilde{\mathbf{f}}$ is composed from two components with frequences ω_1, ω_3 and the noise is composed from two components with frequences ω_2, ω_4, where $0 < \omega_1 \ll \omega_2 \ll \omega_3 \ll \omega_4$. Then, as $t > 0$ increases, the distance $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$ first decreases as ω_4-component of the noise is filtered out, then increases as ω_3-component of the input image attenuates, then again decreases as ω_2-component disappears, etc. Thus the dependence of $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$ on t is not unimodal and has two local minima.

Having discussed the properties of the function $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$, let us investigate the properties of the correlation between the filtering noise and the filtered signal, $\mathrm{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$. Correlation between two vectors (understood as two realizations of a random process) has the geometrical meaning of the angle between the two vectors, as defined through their inner product. In our case (using the assumption $E(\mathbf{u}(0)) = 0$), it is easy to derive that

$$\begin{aligned}&\mathrm{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)) \\ &\quad = \frac{\langle \mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\rangle}{|\mathbf{u}(0) - \mathbf{u}(t)| \cdot |\mathbf{u}(t)|} = \cos\theta \end{aligned} \tag{16}$$

where θ is the angle between vectors $\mathbf{u}(0) - \mathbf{u}(t)$ and $\mathbf{u}(t)$. Then, minimizing the correlation has the geometrical meaning of finding such T that the vectors $\mathbf{u}(0) - \mathbf{u}(T)$ and $\mathbf{u}(T)$ are as orthogonal as possible.

As discussed above, the inner product may be applied to Fourier images,

$$\begin{aligned}\langle \mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t)\rangle &= \langle \boldsymbol{U}(0) - \boldsymbol{U}(t), \boldsymbol{U}(t)\rangle \\ &= \sum_{j,k} (\boldsymbol{U}_{jk}(0) - \boldsymbol{U}_{jk}(t))\overline{\boldsymbol{U}_{jk}(t)}, \end{aligned} \tag{17}$$

where the bar denotes the complex conjugate. For linear diffusion, we apply (15) and derive that the entry with coordinates j, k contributes to the inner product (17) by

$$(\boldsymbol{U}_{jk}(0) - \boldsymbol{U}_{jk}(t))\overline{\boldsymbol{U}_{jk}(t)}$$
$$= (\boldsymbol{U}_{jk}(0) - a_{jk}(t)\boldsymbol{U}_{jk}(0))a_{jk}(t)\overline{\boldsymbol{U}_{jk}(0)}$$
$$= (1 - a_{jk}(t))a_{jk}(t)|\boldsymbol{U}_{jk}(0)|^2$$

which is a nonnegative real number. We conclude that

$$\langle \boldsymbol{U}(0) - \boldsymbol{U}(t), \boldsymbol{U}(t)\rangle \geq 0$$

for all t, and the correlation $\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$ is always nonnegative when the image is filtered using linear diffusion.

In the particular case when the trajectory $\mathbf{u}(t)$ passes through the ideal solution $\tilde{\mathbf{f}}$ for some value of t, the inner product becomes zero and $|\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|$ attains its global minimum exactly at the time T for which $\mathbf{u}(T) = \tilde{\mathbf{f}}$. For real filters which corrupt the signal slightly when removing the noise, none of the possible solutions $\mathbf{u}(t)$ will be identical to the ideal signal. Also, the noise in real data will often exhibit some amount of correlation with the signal to restore. For such situations, we cannot guarantee any theoretical optimality, and the worth of our stopping time estimation can only be evaluated experimentally.

These statements hold analogously for nonlinear diffusion: without a priori restrictions on the nature of the signal, we cannot guarantee that either $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$ or $|\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|$ will have a unique minimum. However, the situation is not so severe in practice. If we filter our data by a filter which respects the expected signal properties, it will first remove most of the noise before harming the signal significantly; in our diffusion experiments, the function $|\mathbf{u}(t) - \tilde{\mathbf{f}}|$ always had a unique minimum and it was always unimodal, i.e. decreasing at the beginning of the diffusion process and increasing later, except for the degenerate cases when there was either no noise or no signal in the input data.

3.2. Experiments

To assess the applicability of the decorrelation criterion in real situation, we perform a set of experiments in which we simulate the image formation process sketched in Fig. 1 by taking some image, either natural of artificially created, and declaring it to be noise-free and represent the ground truth $\tilde{\mathbf{f}}$. We then corrupt the image artificially by adding some noise n, filter the noisy image $\mathbf{u}(0) = \tilde{\mathbf{f}} + n$ using several methods, and

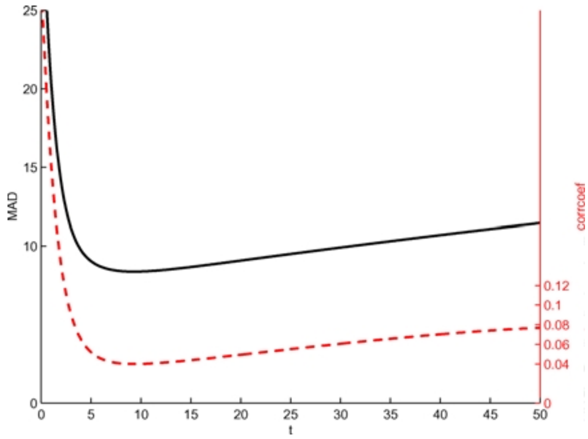

Figure 2. The distance $\text{MAD}(\mathbf{u}(t) - \tilde{\mathbf{f}})$ (solid line) and the correlation coefficient $\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$ (dashed line) developing with the diffusion time. The graphs were measured experimentally on noisy cymbidium data of SNR = 6 filtered using anisotropic NL diffusion.

observe the development of the distance between the filtered signal and the noise-free image, $\text{MAD}(\mathbf{u}(t) - \tilde{\mathbf{f}})$, and the development of the correlation between the filtered signal $\mathbf{u}(t)$ and the 'filtering noise' $\mathbf{u}(0) - \mathbf{u}(t)$.

Ideally, the two graphs will look like those in Fig. 2. You can see that the distance $\text{MAD}(\mathbf{u}(t) - \tilde{\mathbf{f}})$ decreases in the first iterations as the noise is smoothed, and then starts to increase again as also the useful signal begins to disappear under the filtering process. The graph of the correlation coefficient $\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$ exhibits a highly similar behaviour; this similarity lets us hope that we can estimate the stopping time T which optimises the filtering quality (measured by the MAD distance) by locating the minimum of the correlation $\text{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$.

In the experiments below, we will compare several values: T_{opt} will stand for the optimal stopping time which minimizes the distance between the filtered and the ideal data (known in the experimental setup only), and D_{opt} will denote the value of this optimal distance, $D_{\text{opt}} = \text{MAD}(\mathbf{u}(T_{\text{opt}}) - \tilde{\mathbf{f}})$. The symbol T_{corr} will stand for the stopping time estimated using the decorrelation criterion (12), and D_{corr} will represent the filtering residual at such a stopping time. In some cases we also compare the results to the estimation using the method of Weickert (1999) which requires knowledge of the signal-to-noise ratio (SNR) in the input image and iterates the diffusion until the filtered signal is in some distance from the noisy input; the values obtained using this method will be denoted T_{SNR} and D_{SNR}.

The filtering methods we use in our experiments include linear diffusion, nonlinear diffusion filtering, anisotropic NL diffusion filtering, monotonicity-enhancing NL diffusion, and iterated median filtering. In the following paragraphs we review the basics of these methods; more information can be found in the cited references or in the thesis (Mrázek, 2001).

Linear diffusion (Iijima, 1962; Witkin, 1983) is described by the equations

$$\partial_t \mathbf{u} = \Delta \mathbf{u}, \tag{18}$$

$$\mathbf{u}(x, 0) = \mathbf{f}(x). \tag{19}$$

The diffusion starts from the input image $\mathbf{f}$; solution of the heat Eq. (18) at time $t > 0$ is equivalent to a convolution of the input image with a Gaussian kernel,

$$\mathbf{u}(x, t) = (G_{\sqrt{2t}} * \mathbf{f})(x), \tag{20}$$

where

$$G_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-x^T x}{2\sigma^2}\right). \tag{21}$$

The linear diffusion reveals the drawback of smoothing not only the noise, but also blurring the image edges as the time t increases. To make inter-region smoothing preferred to smoothing across edges, Perona and Malik (1990) suggested to slow the diffusion at locations of larger gradient magnitude. In the regularized form due to Catté et al. (1992), the resulting nonlinear (NL) diffusion equation can be written as

$$\partial_t \mathbf{u} = \operatorname{div}(g(|\nabla \mathbf{u}_\sigma|) \cdot \nabla \mathbf{u}), \tag{22}$$

where $\mathbf{u}_\sigma = G_\sigma * \mathbf{u}$. In our experiments, we employ the diffusivity function

$$g(s) = 1 - \exp\left(\frac{c}{(s/\lambda)^m}\right). \tag{23}$$

with $m = 4$ and the constant $c \doteq -2.33667$ determined so that the flux $s \cdot g(s)$ is increasing for $s < \lambda$ and decreasing for $s > \lambda$. This way, the parameter λ serves as a threshold of gradient size: a smaller gradient is diffused, positions of a larger gradient are treated as edges. The value of this parameter can be tuned using some image statistics (Perona and Malik, 1990; Black et al., 1998).

As the isotropic NL diffusion (22) stops the diffusion at image edges, it may leave some noise there. To mitigate this effect, Weickert (1996, 1998) generalized the diffusion equations into the anisotropic case. Then, the diffusion process is not controlled by a scalar diffusivity g, but by a diffusion tensor, and the smoothing in some directions may be preferred. In the so-called 'edge-enhancing' variant, diffusion in the gradient direction is still controlled by a diffusivity function g, but a maximum smoothing is always allowed in the coherence direction (i.e. perpendicular to the direction of maximum gradient, and along edges). In the experiments below, we denote the amount of smoothing in the coherence direction by a parameter φ; its maximum value proposed by Weickert is $\varphi = 1$ (the maximum value of diffusivity g is also 1). When approximating the continuous anisotropic equation by discrete algorithms, this large value of φ may also cause some undesirable blurring in other directions. We will therefore also investigate the possibility to set φ to a smaller value, $\varphi \in [0.05, 0.2]$; this is still sufficient to remove noise from edges while allowing a better preservation of brightness discontinuities.

The diffusion filters mentioned so far tend to attract the input data to a (piecewise) constant model. A generalization of NL diffusion filters for piecewise linear functions was proposed in Mrázek (2002b). This so-called monotonicity-enhancing NL diffusion consists of filtering first directional derivatives of the input data by NL diffusion, and integrating the results.

As the last of the filtering methods tested, iterated median filter replaces in each iteration the value of each pixel by the median of the values in its local neighbourhood. In the experiments below, the median is computed in a 3×3 window.

The images taken for the ideal signal in the experiments are shown in Fig. 3. Some of them were created artificially as discrete versions of constant, piecewise constant or piecewise linear functions in 2D (Fig. 3(a)–(d)), Fig. 3(e) represents a slide from MRI medical dataset, Fig. 3(f)–(g) show photographic images. One complication with the more realistic images 3(e)–(g) is that they are likely to contain some amount of noise; when we declare them to be noise-free, corrupt them with additive noise and measure the distance of the filtered data from them, the results will be biased by the noise originally present.

A first example of the experimental performance of the decorrelation criterion is seen in Figs. 4 and 5, measured on cymbidium data (Fig. 3(f)) with varied amount of additive Gaussian noise combined with an anisotropic NL diffusion filter. The former figure

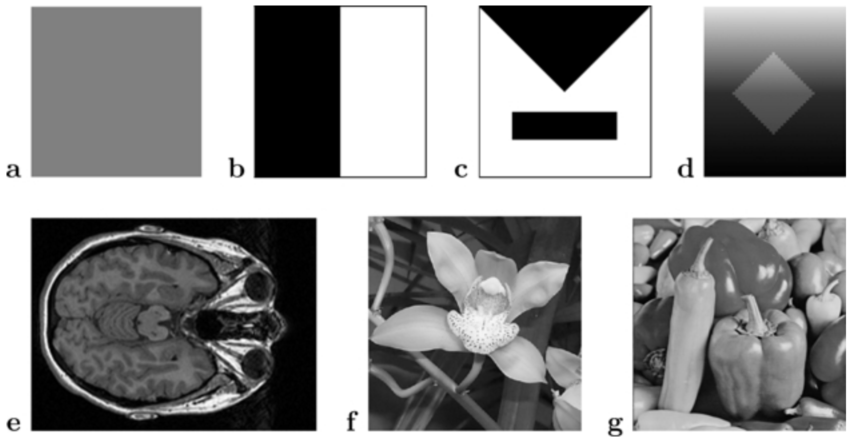

Figure 3. Images used in the experiments.

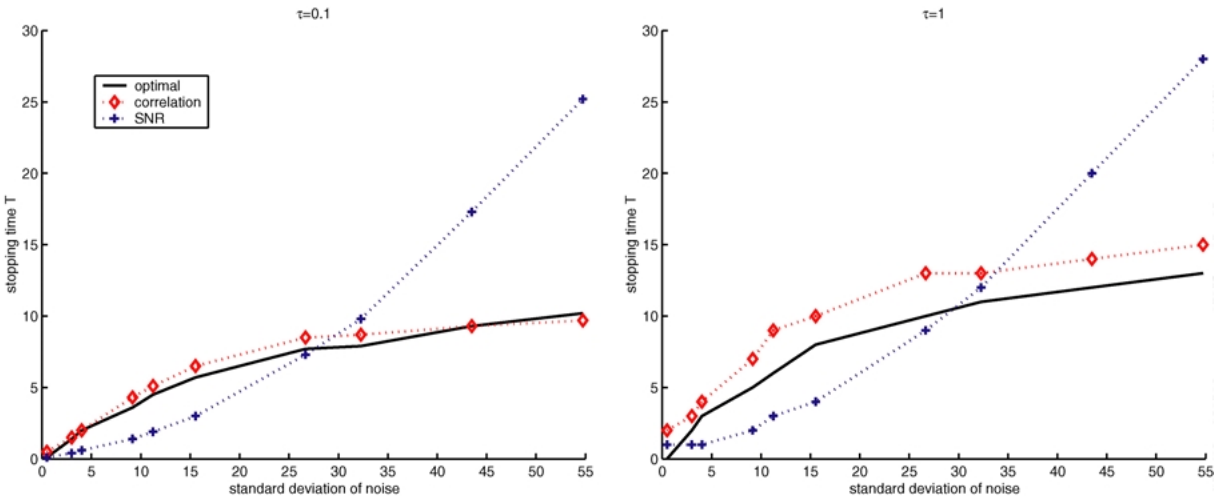

Figure 4. The stopping time T_{SNR} determined by the SNR method (dotted with crosses), and T_{corr} obtained through the covariance minimization (dotted with diamonds) compared to the optimal stopping time T_{opt} (solid line). The graphs are plotted against the standard deviation of noise in the input image; the two figures represent the same measurements for different iteration time-step sizes: $\tau = 0.1$ (left), $\tau = 1$ (right). (Data from Fig. 3(f), anisotropic NL diffusion.)

compares three stopping times: the optimal T_{opt}, T_{SNR}, and our decorrelation-estimated T_{corr}. All alternative stopping times are computed for a series of input images with varied amount of noise present; the standard deviation of noise in the input data is represented by the horizontal axis of the graph. 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 two graphs are plotted for iteration time steps $\tau \in \{0.1, 1\}$.

The actually obtained quality measure $\text{MAD}(\mathbf{u}(T) - \tilde{\mathbf{f}})$ is shown in Fig. 5, this time 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.

3.2.1. Numerical Results. We measured the optimal stopping time T_{opt} and the estimated T_{corr} for all images in Fig. 3, corrupted with varied amount of either additive Gaussian or salt&pepper noise, and filtered using several filters. The values of the ratio $T_{\text{corr}}/T_{\text{opt}}$, averaged for each test image and each filter type across all

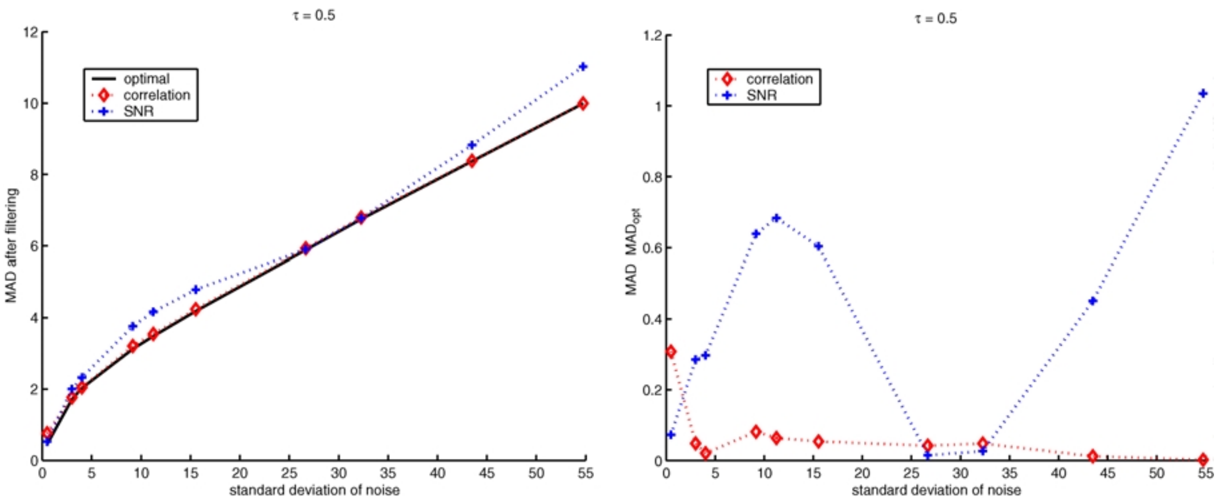

Figure 5. Left: the MAD distance of the filtered data from the ideal noise-free image, MAD($\mathbf{u}(T) - \tilde{\mathbf{f}}$), using the SNR and the correlation-minimization time selection strategies. Right: the difference between the estimated result and the optimal one, MAD($\mathbf{u}(T) - \mathbf{u}(T_{opt})$). (Data from Fig. 3(f), anisotropic NL diffusion.)

noise levels, are given in Table 1. Table 2 presents the corresponding average relative filtering quality

$$\frac{D_{corr}}{D_{opt}} = \frac{\mathrm{MAD}(\mathbf{u}(T_{corr}) - \tilde{\mathbf{f}})}{\mathrm{MAD}(\mathbf{u}(T_{opt}) - \tilde{\mathbf{f}})}. \tag{24}$$

Similar measurements of the ratio T_{corr}/T_{opt}, but presented for each noise level with averaged contribution of individual images, are given in Table 3.

You can see in Tables 1–3 that the estimation of the stopping time T using the decorrelation criterion (12) gives usually good results for linear, isotropic and anisotropic NL diffusion, where the values of T_{corr} lie in most cases in the range $[0.5T_{opt}, 2T_{opt}]$ (more precisely, the estimated values lie in this range in 68, 56, and 71 percent of all experiments for linear, isotropic and anisotropic diffusion, respectively). The estimation results are more reliable when there is a higher amount of noise in the input image.

The errors in the estimation of T lead usually to only small relative decay in the filtering performance compared to the optimal results: the relative error is smaller than 20 percent in 93 percent of experiments with the anisotropic NL diffusion filter, in 86 percent of

Table 1. Ratios between optimal and correlation-estimated stopping times, T_{corr}/T_{opt}, averaged for each combination of test image and filtering method across all noise levels. The noise was additive Gaussian in the range of between 'no noise' and SNR = 1 except for the case f2, for which 0–70 per cent of pixels were corrupted with salt&pepper noise.

	T_{corr}/T_{opt}				
Image	Linear diff.	Iso. NL diff.	Aniso NL diff.	Mono NL diff.	Iter. median
a	1	1	1	1.15	0.07
b	0.71	0.43	0.44	0.8	0.04
c	1.28	0.3	0.65	10.79	0.03
d	0.76	0.72	0.87	1.68	0.12
e	76.32	7.74	2.13	43.01	0.44
f1	1.68	1.49	1.35	21.44	2.53
f2	5.47	4.26	4.09	14.74	0.81
g	1.66	1.38	1.36	5.45	0.16

Table 2. Relative filtering quality $D_{\mathrm{corr}}/D_{\mathrm{opt}}$ computed as a ratio between the filtering residuals $\mathrm{MAD}(\mathbf{u}(T)-\tilde{\mathbf{f}})$ at the correlation-estimated and optimal stopping times T, averaged for each image and filtering method across all noise levels.

	$D_{\mathrm{corr}}/D_{\mathrm{opt}}$				
Image	Linear diff.	Iso. NL diff.	Aniso NL diff.	Mono NL diff.	Iter. median
a	1	1	1	1.01	1.39
b	1.02	1.59	1.19	1.04	1.76
c	1.03	1.52	1.08	2.6	1.74
d	1.02	1.06	1.01	1.08	1.36
e	2.97	1.36	1.18	3.15	1.14
f1	1.07	1.04	1.03	3.18	1.29
f2	1.46	1.25	1.33	3.11	1.38
g	1.07	1.01	1.03	2	1.31

Table 3. Ratios between optimal and correlation-estimated stopping times, $T_{\mathrm{corr}}/T_{\mathrm{opt}}$, averaged for each combination of noise level and filtering method across all test images.

	$T_{\mathrm{corr}}/T_{\mathrm{opt}}$				
SNR	Linear diff.	Iso. NL diff.	Aniso NL diff.	Mono NL diff.	Iter. median
100	1.58	5.55	1.36	18.92	5.35
40	1.61	3.43	1.2	18.67	0.44
20	15.31	2.68	1.17	17.62	0.22
13.33	15.32	2	1.16	15.23	0.19
10	15.29	1.46	1.1	14.97	0.12
6.67	15.35	1.29	1.04	14.76	0.11
5	15.17	1.13	1.08	13.6	0.08
3.33	15.28	1.15	1.05	8.28	0.07
2.5	7.99	1.21	1.08	7.5	0.07
2	7.81	1.14	1.06	7.25	0.08
1.67	7.75	1.13	1.08	6.58	0.07
1.25	7.58	1.08	1.06	5.13	0.09
1	7.5	0.99	1.07	4.68	0.08

experiments with the linear diffusion, and in 71 percent of experiments with the isotropic NL diffusion filtering. Note also that this higher error with isotropic NL diffusion is only relative to the optimal values obtained by that filter, the absolute results of isotropic NL filter with estimated stopping time (as measured by the distance $\mathrm{MAD}(\mathbf{u}(T_{\mathrm{corr}})-\tilde{\mathbf{f}})$) outperform other methods in most cases; see the research report (Mrázek, 2002a) for details.

The estimation works well with linear diffusion except the MRI data (Fig. 3(e)). On the other hand, the minimum of the signal-noise correlation does not lead to near-optimal stopping times with monotonicity-enhancing NL diffusion and with iterated median filters, perhaps with the exception of the data suited directly for these types of filters (i.e. the piecewise linear function, 3d, and the salt&pepper noise in experiment f2, respectively). This worse performance may be caused by the fact that unlike classical diffusion filters which create information-reducing scale-spaces and observe the maximum-minimum principle, monotonicity-enhancing diffusion and iterated median filter often form a solution for which the correlation $\mathrm{corr}(\mathbf{u}(0)-\mathbf{u}(t),\mathbf{u}(t))$ changes its sign into negative values. Also, the median filtering does not preserve the average grey level of the image, which

Figure 6. Noisy input image for the experiment in Fig. 7.

was one of the theoretical assumptions on the filtering methods.

3.2.2. Some Filtered Images. In this section, we show some images obtained using diffusion filtering combined with autonomous stopping time selection using the decorrelation criterion (12).

The first example 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. 6: the noisy image was obtained by adding noise of uniform distribution in the range $[-255, 255]$ to the ideal input (Fig. 3(c)), and by restricting the noisy values into the interval $[0, 255]$.

In Fig. 7, 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 (12). 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. 8.

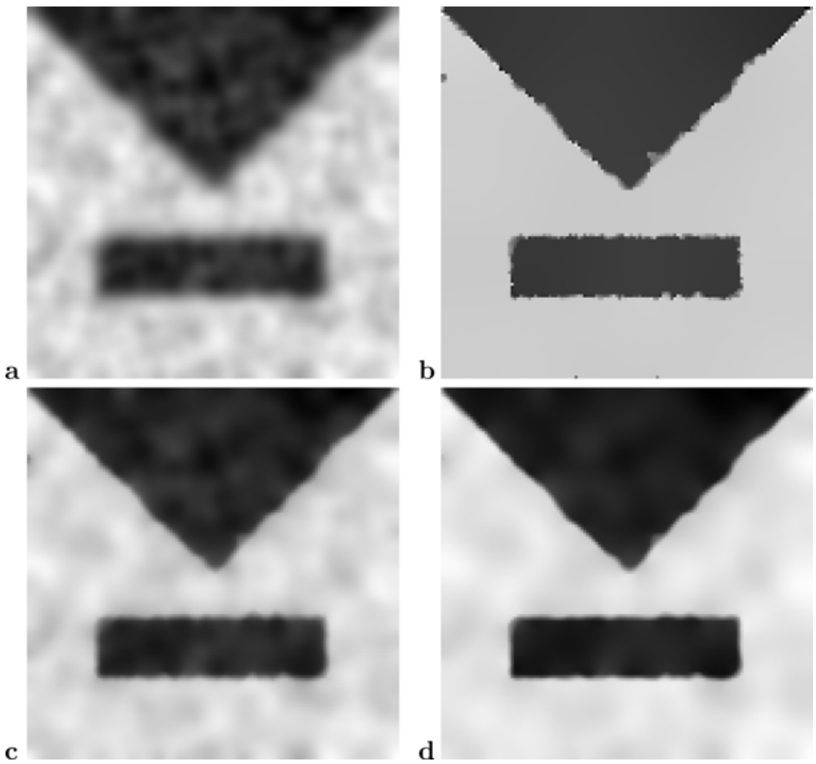

Figure 7. Comparing the different diffusion algorithms on the noisy data of Fig. 6, all with the stopping time selected autonomously by minimizing the criterion (11): (a) linear diffusion, $T = 3.8$; (b) isotropic nonlinear diffusion, $T = 125$; (c) anisotropic NL diffusion, $\varphi = 1$, $T = 15$; (d) anisotropic NL diffusion, $\varphi = 0.2$, $T = 32$.

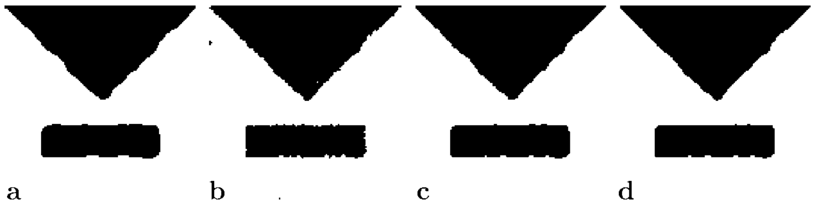

Figure 8. Thresholded versions of the images in Fig. 7.

The stopping criterion was designed to minimize the distance between the ideal and filtered function. If visual quality was the goal to be achieved, we would probably stop the diffusion later, especially as linear diffusion (Fig. 7(a)) and the Weickert's edge-enhancing anisotropic diffusion (Weickert, 1998) with maximum amount of diffusion in the coherence direction ($\varphi = 1$, Fig. 7(c)) are concerned. We find however that the MAD distance and visual quality are in a good agreement in Fig. 7(d) which represents the result of the edge-enhancing diffusion with a smaller amount of diffusion in the coherence direction, $\varphi = 0.2$.

In the second example, noise of normal distribution was added to the image of a cymbidium flower (Fig. 3(f)), and the noisy image was filtered using anisotropic NL diffusion combined with our decorrelation method to determine the optimal T. The results for two levels of noise can be seen in Fig. 9.

Next, we present an experiment with piecewise linear data. The ideal signal (the image 3d visualized as a surface in 3D, where the original grey level is converted to elevation) and the noisy input are drawn at the top of Fig. 10. Such an increasing surface with discontinuities is not easily filtered using classical diffusion filters

Figure 9. The cymbidium experiment. Left column (top to bottom): input images of SNR $=$ 14.87 and 6.17. Right column: corresponding images filtered by anisotropic NL diffusion. The stopping time T was chosen autonomously using the decorrelation criterion.

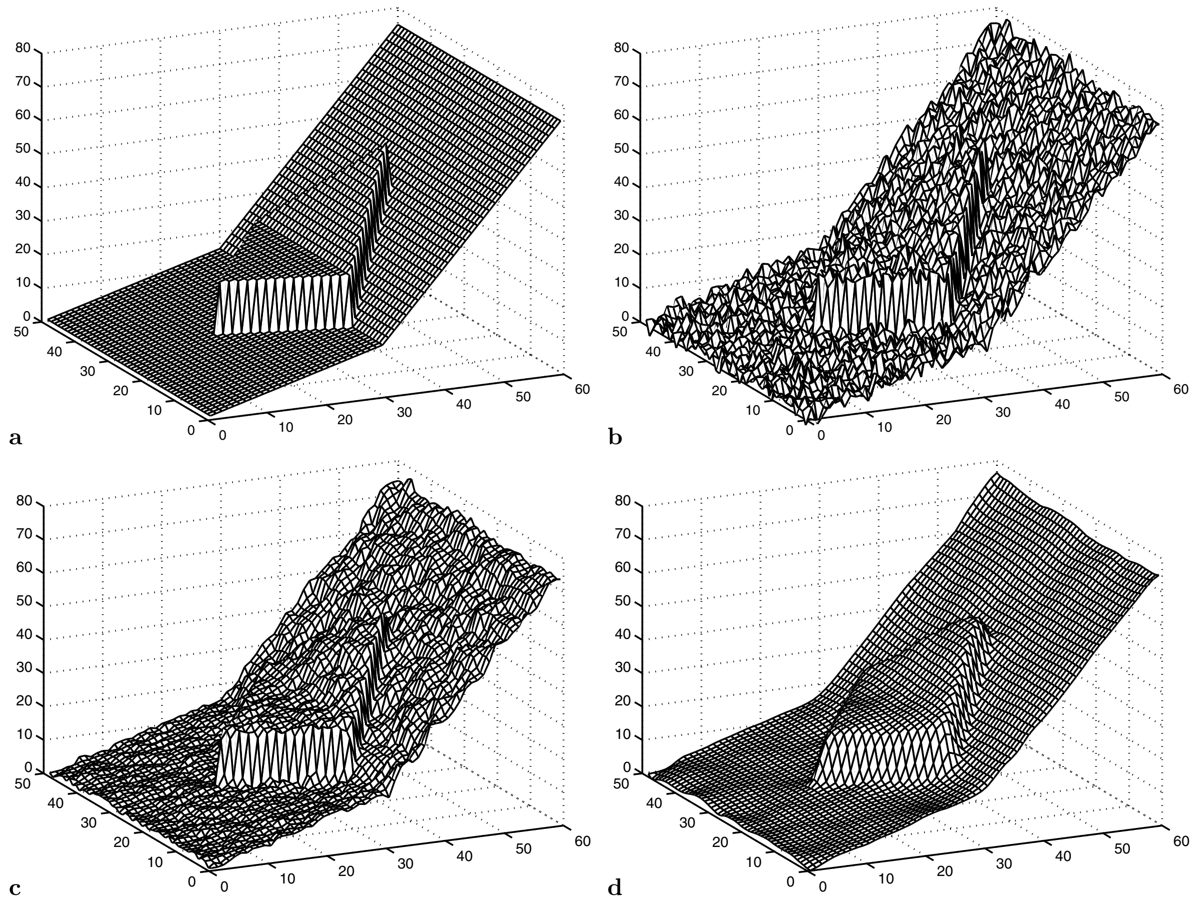

Figure 10. Experiment with two-dimensional data. (a) Ideal data. (b) Noisy input. (c) Classical anisotropic NL diffusion, $T = 4$. (d) Anisotropic NL diffusion for monotonicity-enhancement, $T = 10$. In (c) and (d), the stopping time was determined autonomously using the decorrelation criterion.

which are based on the assumption that the data to be filtered are piecewise constant. The image gradient is larger on the sloped surface, and it is rather difficult to tune the parameters of the classical nonlinear diffusion to smooth the noise there without removing the signal discontinuities. Piecewise linear or piecewise increasing data can be recovered better using specialized monotonicity-enhancing diffusion filters, see (Mrázek, 2001, 2002b).

The results of two anisotropic NL diffusion filters with autonomous selection of T using the decorrelation criterion can be seen in the bottom row of Fig. 10 (the classical one on the left and the monotonicity-enhancing on the right). While the classical filter leads to a rather rough result, trying to approximate the surface with piecewise constant patches, the monotonicity-enhancing procedure is better suited for this kind of data, combines well with the time-selection procedure, and provides a good estimate of the ideal signal.

3.3. *Further Applications*

So far, we have been addressing the problem of selecting a good stopping time T from a sequence of iterations. However, the correlation between the filtered signal and the filtering noise $\mathrm{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))$ carries more information on the filtering process and can be exploited further: to estimate the filtering quality, to select parameters of the filtering process, or choose between several alternative filtering methods.

Let us return for a moment to Fig. 2. At the beginning of the diffusion filtering, the correlation coefficient declines fast until it reaches its minimum. 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 $\mathrm{corr}(\mathbf{u}(0)-\mathbf{u}(t), \mathbf{u}(t))$ 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.

At its minimum, the term $|\mathrm{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|$ measures the residual correlation between the filtered signal and the filtering noise. It can be also understood as a measure of the filtering quality obtained by a particular filter with a given set of parameters; hopefully, a smaller residual correlation corresponds to a better filtering quality (the correlation is zero for the ideal filter). Then, the minimum of $|\mathrm{corr}(\mathbf{u}(0)-\mathbf{u}(t), \mathbf{u}(t))|$ can be used to compare the filtering results of different filters or of one filter with different parameters, and may help us choose the best filter for the given input data. Some initial experiments in this direction are presented in the report (Mrázek, 2002a).

4. 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 minimized. This method, which we call *decorrelation criterion* is very general, being based only on the assumptions that the noise and the signal in the input image are uncorrelated, that neither the additive noise nor the filtering procedure alter the average gray value, and that the filtering method is suitable for that given type of signal. The estimation results are more reliable for filters for which the sign of $\mathrm{corr}(\mathbf{u}(0)-\mathbf{u}(t), \mathbf{u}(t))$ does not change. No knowledge on the variance of the noise, and no training images are needed to tune any parameters of the method.

We have analysed the theoretical properties of the decorrelation criterion, and demonstrated that the results of both the filtering procedure and of our stopping time estimation depend on the signal and noise properties, and on the filtering method used; in general, neither the filtering quality measured by the distance from the ideal signal, nor the values of the correlation $|\mathrm{corr}(\mathbf{u}(0)-\mathbf{u}(t), \mathbf{u}(t))|$ are guaranteed to exhibit a single minimum. However, the experiments suggest that in practical situations with diffusion filters the first local minimum of the correlation is also the global one (i.e. we can easily locate it if we continue the diffusion iterations until the correlation starts to increase), and the decorrelation criterion provides a good estimate of the optimal stopping time for a wide range of noise levels and filtering parameters. In some cases, the new criterion even outperforms other time selection strategies which use more a priori information e.g. about the noise variance; still, if more information is available, we would suggest to use it, compute several stopping time estimates using different methods, and compare the results to improve reliability of the estimation.

The minimum of the correlation $|\mathrm{corr}(\mathbf{u}(0) - \mathbf{u}(t), \mathbf{u}(t))|$ can be also used to evaluate the filtering performance of the filter which created the sequence $\mathbf{u}(t)$. This measure can then be used to adapt other parameters of the diffusion, or choose one of several filters which is most suitable for the input data. We have mentioned some initial ideas in Section 3.3, but many questions in this direction still have to be answered by further research.

Appendix: Statistical Definitions

Let us review the notation and statistical definitions used in the paper (see e.g. Papoulis (1990)). 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

$$\bar{x} = E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

We define the *variance* of a signal x as

$$\mathrm{var}(x) = E[(x - \bar{x})^2].$$

The *covariance* of two vectors x, y is given by

$$\mathrm{cov}(x, y) = E[(x - \bar{x}) \cdot (y - \bar{y})].$$

The normalized form of the covariance is called the *correlation coefficient*,

$$\mathrm{corr}(x, y) = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x) \cdot \mathrm{var}(y)}}.$$

To evaluate the filtering quality, we need the notion of distance between two images. The Euclidean distance induced by the 2-norm seems most suited for theoretical analysis. In our experiments, we measure the distance of two images by the *mean absolute deviation*, $\mathrm{MAD}(x - y) = E(|x - y|)$, which is equivalent to the 1-norm of the difference vector normalized by the number of pixels in the image. Compared to the Euclidean distance, the MAD distance is less sensitive to outliers.

Acknowledgments

The first author was supported by the Czech Ministry of Education under project LN00B096; the second author was supported by the Czech Ministry of Education under project MSM 212300013. The final revisions were completed after P. Mrázek joined the Mathematical Image Analysis group at the Saarland University, Germany, where his research is supported by DFG-Schwerpunktprogramm 1114 'Mathematical methods for time series analysis and digital image processing'.

The authors are grateful to many people who contributed with their comments on the topics covered in this paper: Eduard Krajník, our colleagues at the Center for Machine Perception of the Czech Technical University, Joachim Weickert, and the participants of the Scale-Space'01 conference in Vancouver.

References

Bertero, M. and Boccacci, P. 1998. *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, Bristol, UK.

Black, M.J., Sapiro, G., Marimont, D., and Heeger, D. 1998. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432.

Capuzzo Dolcetta, I. and Ferretti, R. 2001. Optimal stopping time formulation of adaptive image filtering. *Applied Mathematics and Optimization*, 43:245–258.

Catté, F., Lions, P.-L., Morel, J.-M., and Coll, T. 1992. Image selective smoothing and edge-detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193.

Iijima, T. 1962. Basic theory on normalization of a pattern. *Bulletin of Electrical Laboratory*, 26:368–388 (in Japanese).

Mrázek, P. 2001. Nonlinear diffusion for image filtering and monotonicity enhancement. Ph.D. Thesis, Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic. Available at `ftp://cmp.felk.cvut.cz/pub/cmp/articles/mrazek/Mrazek-phd01.pdf`.

Mrázek, P. 2002a. Decorrelation criterion to select diffusion stopping time: Experimental evaluation. Research Report K333-13/02, CTU-CMP-2002-01, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic. Available at `ftp://cmp.felk.cvut.cz/pub/cmp/articles/mrazek/Mrazek-TR-2002-01.pdf`.

Mrázek, P. 2002b. Monotonicity enhancing nonlinear diffusion. *Journal of Visual Communication and Image Representation*, 13(1/2):313–323.

Papoulis, A. 1990. *Probability and Statistics*. Prentice-Hall: Englewood Cliffs, NJ.

Perona, P. and Malik, J. 1990. Scale-space and edge-detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

Rudin, L.I., Osher, S., and Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica* D, 60:259–268.

Sporring, J. and Weickert, J. 1999. Information measures in scale-spaces. *IEEE Transactions on Information Theory*, 45:1051–1058.

Weickert, J. 1996. Theoretical foundations of anisotropic diffusion in image processing. *Computing (Suppl. 11)*, pp. 221–236.

Weickert, J. 1998. *Anisotropic Diffusion in Image Processing*. European Consortium for Mathematics in Industry (B.G. Teubner, Stuttgart).

Weickert, J. 1999. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17:201–212.

Witkin, A.P. 1983. Scale space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, A. Bundy (Ed.), Karlsruhe, Germany. William Kaufmann, pp. 1019–1023.