Pre-process for segmentation task with nonlinear diffusion filters

Javier Sanguino Carlos Platero
Olga Velasco

Health Science Technology Group, Technical University of Madrid,
Ronda de Valencia 3, 28012, Madrid, Spain.

Manuscript originally written in January 2017. Submitted to arXiv, 2026.

Abstract

This paper deals with the case of using nonlinear diffusion filters to obtain piecewise constant images as a previous process for segmentation techniques.

We first show an intrinsic formulation for the nonlinear diffusion equation to provide some design conditions on the diffusion filters. According to this theoretical framework, we propose a new family of diffusivities; they are obtained from nonlinear diffusion techniques and are related with backward diffusion. Their goal is to split the image in closed contours with a homogenized grey intensity inside and with no blurred edges.

We also prove that our filters satisfy the well-posedness semi-discrete and full discrete scale-space requirements. This shows that by using semi-implicit schemes, a forward nonlinear diffusion equation is solved, instead of a backward nonlinear diffusion equation, connecting with an edge-preserving process. Under the conditions established for the diffusivity and using a stopping criterion for the diffusion time, we get piecewise constant images with a low computational effort.

Finally, we test our filter with real images and we illustrate the effects of our diffusivity function as a method to get piecewise constant images.

The code is available at `https://github.com/cplatero/NonlinearDiffusion`.

Keywords: Nonlinear diffusion, PDE, Cartoon images, Segmentation
MSC classes: 68U10 (Image processing), 68T45 (Machine vision and scene understanding), 65M06 (Finite difference methods)

1 Introduction

A typical application of the Computer Vision is the segmentation process, where initial images show noise, regions without homogeneity in their intensities, weak edges and small artifacts inside of regions of interest.

However, segmentation techniques are usually based on the fact that the regions are piecewise constant [10] and the edges have high slopes. Hence segmentation approaches need that the images have to be filtered previously in order to homogenize the regions and enhance the edges.

Therefore, given an observed image u_0 we are interested in finding another image u, *close* to u_0, formed by homogeneous regions and with sharp boundaries. The *piecewise constant images* or *cartoon images*, that we are interested to obtain, should be composed by regions delimited by closed and sharped contours with a constant grey value inside.

The cartoon images arise in some models in image analysis where a relation is assumed between an observed image u_0 and the cartoon component u as follows: $u_0 = u + v$ where v is noise or small scale repeated detail (texture). In this case, both types of patterns (additive noise or texture) can be modeled by oscillatory functions taking both positive and negative values and zero mean [24]. These models yield to a total variation minimization problem where $u \in BV$ is a function of Bounded Variation, while v belongs to G, a space of oscillating functions which is the dual of the closure in $BV(\mathbb{R}^2)$ in the Schwartz class [24]. In some contexts the v component is important, especially if we are interested in image texture [3, 4, 7, 24, 40]. On the other hand, some models in a context of image recovery or classical image denoising suggest that the v component models the noise in a total variation minimization framework [2, 9, 33, 39].

For image segmentation, the Mumford-Shah model [26] is one of the first image decompositions in texture and cartoon components. Furthermore, it is also known the connection between the Mumford and Shah model and the Perona-Malik evolution equations [29] as it is shown in [20]. We use this fact as a motivation to get a nonlinear diffusion framework in order to obtain a kind of cartoon image from the initial image. Thus, in this work we are

interested in extracting the cartoon component u from u_0 eliminating the component v that contains noise and some texture of the initial image. Our approach is not based on a total variation minimization model, where the functional has two terms: one measure the fidelity to the data and the other is a smoothing term [1, 2, 33]. We only consider a smoothing term, with a nonconvex functional of type

$$\Psi(u) = \int_{\Omega} \rho\left(\|\nabla u\|\right)\, d\boldsymbol{x} \tag{1}$$

We will study, for continuous images the model described by Perona and Malik [29]. Thus, we begin from the definition of *scalar potential diffusivity function*

$$\rho(s) = \frac{1}{2}\int_0^{s^2} g(t)\, dt \tag{2}$$

where $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ is initially a smooth enough positive function called *diffusivity*. Hence we consider the functional given by (1) and we minimize it by using the descent gradient method to get the following Euler-Lagrange equation

$$\begin{cases} \partial_t u = \text{div}\,\left(g(\|\nabla u\|^2)\,\nabla u\right) & \text{if} \quad (\boldsymbol{x}, t) \in \Omega \times (0, T) \\ u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) & \text{if} \quad \boldsymbol{x} \in \Omega \\ \partial_{\boldsymbol{n}} u = 0 & \text{if} \quad (\boldsymbol{x}, t) \in \partial\Omega \times (0, T] \end{cases} \tag{3}$$

Given an initial image $u_0 : \Omega \to \mathbb{R}$ defined over a bounded domain $\Omega \subset \mathbb{R}^s$ (with $s = 1, 2$ or 3) we get an image $u(\boldsymbol{x}, T)$ as the solution to a nonlinear diffusion equation with initial and Neumann boundary conditions.

The aim of this paper is to propose a diffusivity function with an edge-preserving behaviour that allows us to implement a criterion for establishing a stopping time of the diffusion process. We present a new and a robust procedure to split the original image into piecewise constant regions. In this regard, Black *et al* [5] established that nonlinear diffusion can be seen as a robust estimation procedure getting a piecewise constant image from an observed image. So, our proposal is to present a diffusivity function in order to obtain piecewise constant images using nonlinear diffusion techniques.

It is clear that the piecewise constant images should be related with the edge enhancement or at least an edge-preserving process, so we first propose a theoretical framework for getting these kind of diffusivities.

In this regards, these behaviors are directly related to the non convexity of ρ given by (2) [1, 2]. In this case, the problem is ill-posed and might have no solution, and one cannot prove any convergence result. One way to transform it to a *well-posed* problem is proposed by Catte *et al* [8]. In this case the image u_0 is convolved with a Gaussian mask of standard deviation σ to obtain u_σ so they get a pre-smoothing mechanism. This pre-smoothing mechanism can be a drawback in an edge enhancement process. Nevertheless, discretizations may have a regularizing effect to Perona-Malik 'ill-posed' problem as it is shown by Weickert [45]. Using the *Method of Lines* (MOL), we arrive to a semi-discretization of the problem (3) and we show that our proposed diffusivity satisfies the criterion for semi-discrete diffusion scale-spaces framework [43]; this is sufficient to transform the former partial differential equation (3) in a well-posed system of ordinary differential equations [6]. However, these discrete and iterative solutions tend to the average grey value of the image. Therefore, it is necessary that we introduce a criterion to stop the diffusion process to facilitate the access to the appropriate images.

Forward-and-Backward (FAB) diffusion model proposed by Gilboa et al [17] give us another research area [41, 42, 47] where its diffusivities are positive in certain regions and negative in others in order to combine smoothing and sharpening actions. As a consequence, it seeks an enhancement process instead of an edge-preserving behaviour. This model has some difficulties to satisfy the well-posedness semi-discrete and full discrete scale-space requirements [47]. For these reasons, in this work, we have chosen to use an edge-preserving process where we have the possibility of using semi-implicit time discretizations that allows us to use a larger time steps with the idea of reducing the computational effort.

The outline of the paper is as follows. First, based upon the Perona and Malik model (3), we present an intrinsic formulation that can be used to set the design conditions for edge enhancement diffusivity functions for 2 and 3 dimension, in order to get a theoretical framework that we will connect with a numerical scheme related with an edge-preserving process. It is also shown that the Perona-Malik function [29] and some others diffusivities are particular cases of these design conditions. In section 3, we propose a diffusivity function with the aim to obtain a piecewise constant image from an observed image and with the feature of providing a stopping time for the diffusion process. In section 4, we show that our diffusivity achieves the necessary conditions for the existence and uniqueness of solutions using the

scale-space framework established by Weickert and Benhamouda [45]. For simplicity, we begin applying the Method of Lines in 1D. Then we proceed with time discretization. We study the implicit scheme from the Backward Euler method and how it leads to a nonlinear algebraic problem. The important point to note here is the singularity of the iteration matrix. We analyze the Jacobian of the system of the differential equation: it is a function of two matrices, one of them containing the effects of the backward diffusion. To avoid the singularities of the iteration matrix, we propose to use the *tangential method* [28, 31] which leads to *Picard iteration* and we show the condition for convergence. We realize that the Picard scheme agrees with the semi-implicit scheme for one dimension. It also allows to point out that in fact we do not do backward diffusion but forward diffusion with a positive decreasing diffusivity function that slows down the diffusion connecting with an edge-preserving process. In Section 5 we extend straightforwardly the results obtained in 1D to 2D and 3D; this leads to the *Additive Operator Splitting* (AOS) method [43, 46]. Finally, in the last section we validate our diffusivity with a semi-implicit scheme for real images in 2D and abdominal Computer Tomography (CT) scans.

2 Condition for edge enhancing in a nonlinear diffusion process

In this section, we establish the condition that the diffusivity g should satisfy in order to get an edge enhancement process with equation (3).

2.1 Intrinsic formulation for nonlinear diffusivity

According to the expression of ρ given in (2) we get: $\rho'(s) = g(s^2)\, s$. Then we define the *diffusion flow* as:

$$\boldsymbol{F} = g(\|\nabla u\|^2)\, \nabla u \tag{4}$$

and $\boldsymbol{\eta} = \nabla u / \|\nabla u\|$ with $\|\nabla u\| \neq 0$. Using $\boldsymbol{\eta}$ and adding orthogonal and unitary vectors we make positive oriented systems $\widehat{B} = \{\boldsymbol{\xi}, \boldsymbol{\eta}\}$ in $\mathbb{R}^2$ and $\widehat{B} = \{\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta}\}$ for $\mathbb{R}^3$.

Now, if $\boldsymbol{I}$ is the identity tensor, 'Tr' is the Trace operator and '**:**' is the

tensor scalar product ($\mathbf{A} : \mathbf{B} = Trace(\mathbf{A}^t \cdot \mathbf{B})$, see Gurtin [18]) we can write

$$\begin{aligned} Tr(\nabla \boldsymbol{F}) = \boldsymbol{I} : \nabla \boldsymbol{F} &= \text{div } \big(g(\|\nabla u\|^2)\, \nabla u\big) \\ &= \underbrace{\nabla g(\|\nabla u\|^2) \cdot \nabla u}_{\text{Transport}} + \underbrace{g(\|\nabla u\|^2)\, \Delta u}_{\text{Diffusion}} \end{aligned}$$

However, we are interested in a intrinsic expression that shows, how the diffusivity g affects the differential operator in equation (3). This is done as follows:

Proposition 1. *If u and g are smooth enough then*

$$\text{div } \big(g(\|\nabla u\|^2)\, \nabla u\big) = \Big[g(\|\nabla u\|^2)\boldsymbol{I} + 2\, g'(\|\nabla u\|^2)\, \nabla u \otimes \nabla u\Big] : \nabla(\nabla u) \quad (5)$$

where the size of identity tensor depends on space dimension.

Proof. We will use some basic tensor properties [18]. From (4), we get

$$\nabla \boldsymbol{F} = \nabla u \otimes \nabla g(\|\nabla u\|^2) + g(\|\nabla u\|^2)\, \nabla(\nabla u)$$

Using the properties of the *Trace* operator, we get

$$Tr(\nabla \boldsymbol{F}) = Tr(\, \nabla u \otimes \nabla g(\|\nabla u\|^2)\,) + Tr(\, g(\|\nabla u\|^2)\, \nabla(\nabla u)\,)$$

and we know that

$$\nabla g(\|\nabla u\|^2) = 2g'(\|\nabla u\|^2)\nabla(\nabla u)[\nabla u]$$

and recalling the symmetry of $\nabla(\nabla u)$, we have

$$\begin{aligned} \nabla u \otimes \nabla g(\|\nabla u\|^2) &= \nabla u \otimes 2g'(\|\nabla u\|^2)\nabla(\nabla u)[\nabla u] \\ &= 2\, g'(\|\nabla u\|^2)\, \{\nabla u \otimes \nabla(\nabla u)[\nabla u]\} \\ &= 2g'(\|\nabla u\|^2)\, (\nabla u \otimes \nabla u)\, \nabla(\nabla u). \end{aligned}$$

For the last equality, we have followed [18, p. 9; 6(b)]. Now taking traces in the last result we get

$$\begin{aligned} Tr(\, \nabla u \otimes \nabla g(\|\nabla u\|^2)\,) &= 2g'(\|\nabla u\|^2) Tr(\, (\nabla u \otimes \nabla u)\, \nabla(\nabla u)\,) \\ &= 2g'(\|\nabla u\|^2)\, (\nabla u \otimes \nabla u) : \nabla(\nabla u) \end{aligned}$$

On the other hand,

$$Tr(\, g(\|\nabla u\|^2)\, \nabla(\nabla u)\,) = g(\|\nabla u\|^2)\, \boldsymbol{I} : \nabla(\nabla u)$$

Therefore

$$Tr(\nabla \boldsymbol{F}) = \Big[g(\|\nabla u\|^2)\boldsymbol{I} + 2\, g'(\|\nabla u\|^2)\, \nabla u \otimes \nabla u \Big] : \nabla(\nabla u)$$

so we arrive to the intrinsic expression (5).

□ □

We note that the Laplacian operator Δ is intrinsic respect any basis. For example, it is easy to check that for 3D, we have

$$\begin{aligned} \Delta_3 u &= (\boldsymbol{\xi} \otimes \boldsymbol{\xi} + \boldsymbol{\zeta} \otimes \boldsymbol{\zeta} + \boldsymbol{\eta} \otimes \boldsymbol{\eta}) : \nabla(\nabla u)_{\widehat{B}} \\ &= (\boldsymbol{i} \otimes \boldsymbol{i} + \boldsymbol{j} \otimes \boldsymbol{j} + \boldsymbol{k} \otimes \boldsymbol{k}) : \nabla(\nabla u)_{B} \end{aligned} \tag{6}$$

where $\widehat{B}$ is the orthonormal basis previously mentioned (see equation (4)). Then, if we use the orthonormal basis $\widehat{B}$ above, we obtain

$$\partial_t\, u(\boldsymbol{x}, t) = \big[g(\|\nabla u\|^2)\boldsymbol{I} + 2\, g'(\|\nabla u\|^2)\|\nabla u\|^2\, \boldsymbol{\eta} \otimes \boldsymbol{\eta} \big] : \nabla(\nabla u)_{\widehat{B}} \tag{7}$$

and we get,

$$\begin{aligned} \partial_t\, u(\boldsymbol{x}, t) = {} & g(\|\nabla u\|^2)\, u_{\boldsymbol{\xi}\boldsymbol{\xi}} + g(\|\nabla u\|^2)\, u_{\boldsymbol{\zeta}\boldsymbol{\zeta}} \\ & + \Big[g(\|\nabla\, u\|^2) + 2\, g'(\|\nabla u\|^2)\|\nabla u\|^2 \Big]\, u_{\boldsymbol{\eta}\boldsymbol{\eta}}. \end{aligned}$$

where $u_{\boldsymbol{\xi}\boldsymbol{\xi}}, u_{\boldsymbol{\zeta}\boldsymbol{\zeta}}, u_{\boldsymbol{\xi}\boldsymbol{\xi}}$ are the elements of the main diagonal of $\nabla(\nabla u)_{\widehat{B}}$. If ρ is given by (2) with $s = \|\nabla u\|$ then:

$$\rho'\,(\|\nabla u\|) = g\left(\|\nabla u\|^2\right)\, \|\nabla u\| \tag{8a}$$

$$\rho''\,(\|\nabla u\|) = g\left(\|\nabla\, u\|^2\right) + 2\, g'\left(\|\nabla u\|^2\right)\, \|\nabla u\|^2 \tag{8b}$$

then, we have another expression for $\partial_t\, u(\boldsymbol{x}, t)$

$$\partial_t\, u(\boldsymbol{x}, t) = \frac{\rho'(\|\nabla u\|)}{\|\nabla u\|}\, u_{\boldsymbol{\xi}\boldsymbol{\xi}} + \frac{\rho'(\|\nabla u\|)}{\|\nabla u\|}\, u_{\boldsymbol{\zeta}\boldsymbol{\zeta}} + \rho''(\|\nabla u\|)\, u_{\boldsymbol{\eta}\boldsymbol{\eta}} \tag{9}$$

with $\|\nabla u\| \neq 0$. Analogous results can be obtained for $2D$ and $1D$ as particular cases.

An interesting insight into nonlinear diffusion filtering may also be gained by looking at its relation with energy minimization. If we consider the functional given by (1), with ρ given by (2) with $s = \|\nabla u\|$, it can be deduced that the second variation is formally equal to:

$$\delta^2\Psi(u) \equiv 2\,g'(\|\nabla u\|^2)(\nabla u \otimes \nabla u) + g(\|\nabla u\|^2)\boldsymbol{I} \tag{10}$$

which it is related with the first operator of the right member of the equation (5). If ρ is strictly convex, the quadratic form (10) is positive, hence the potential Ψ is convex, so it has exactly one minimum. Nevertheless, we are interested in the case when ρ is not convex. In fact, to get images that are piecewise constant and to obtain clear and sharp edges it is necessary to increase the contrast between different regions. That is why this process should be related to an enhancement process. Thus, the condition imposed (see Aubert and Kornprobst [1, p. 122])

$$\rho''(\|\nabla u\|) = g(\|\nabla u\|^2) + 2\,g'(\|\nabla u\|^2)\|\nabla u\|^2 < 0 \tag{11}$$

As mentioned before, this makes ρ not convex. Therefore it is not guaranteed the existence of solutions of equation (3).

2.2 Conditions for diffusivity functions

We propose to use the result in (11) as the condition for edge enhancement diffusivity functions [1]. To do this, observe that (11) can be written as

$$\frac{2\,g'(s^2)}{g(s^2)}\,s^2 < -1 \iff \frac{f'(s)}{f(s)}\,s < -1, \tag{12}$$

where $f(s) = g(s^2)$, which give us conditions that should be verified by the enhancement diffusivity functions. One way to impose this condition is to get $f(s)$ as the solution of the differential equation [1]

$$\frac{f'(s)}{f(s)}\,s = -p < -1$$

where $p > 1$ and $s \neq 0$. The solution is easily obtained: $f(s) = \frac{1}{s^p}$ and thus,

$$g(s^2) = \frac{1}{(s^2)^{p/2}}. \tag{13}$$

[1] Another possibility is to consider $g(s^2) = -\frac{s^2}{\tau^2}$ that leads to the Perona-Malik function $f(s) = e^{-s^2/2\tau^2}$

Thus, based on (13), we define the function called *enhancement diffusivity function*:

$$g(r) = \frac{1}{r^{p/2}} \tag{14}$$

This kind of diffusivity has been already studied for a continuous model in the image denoising context using techniques of the porous media equations [35].

3 Pre-processing function for segmentation task

For the diffusivity, given by (13), it is easy to check that

$$\rho''(s) = g(s^2) + 2 \cdot g'(s^2) \cdot s^2 = \frac{1}{(s^2)^{p/2}} (1-p) < 0 \, \forall \, p > 1.$$

We now focus our attention in this kind of functions because with them we get backward diffusion if $p > 1$. A regularized version of this diffusivity is

$$g(s^2) = \frac{1}{(s^2 + \varepsilon^2)^{p/2}}, \tag{15}$$

with $\varepsilon > 0$ a fix and small constant introduced for numerical purposes. A parameter as ε has been used before in the context of denoising process, when the goal was to reduce smoothing across the edges yielding to the Edge Enhancing Flow [6]; furthermore, with $p = 2$, it leads to the Balanced Forward-Backward Diffusion [21]. It is clear that if ε is small, the function given in (15) can take large values when $\|\nabla u\| \approx 0$. Hence, this regularization presents unreliable results for large values of time steps Δt when it is implemented.

The goal is to get functions with a behaviour similar to the *bounded step function* $B(s)$ (see figure 1). In this case, we set linear diffusion for slopes less than μ_0 and no diffusion otherwise. This will allow us to establish a stopping criterion based on the *setting time* for the linear diffusion. Therefore we can select the piecewise constant regions by changing the parameter μ_0. The disadvantage of this function is that it is not continuous.

For this purpose, using the function $g(t)$ in (2) and its derivative, the enhancement diffusivity function given by (14) and the bounded step function

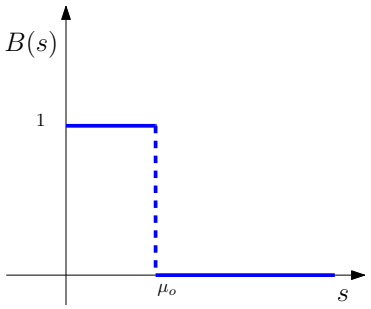


Figure 1: Bounded step function

$B(s)$, we propose the following diffusivity function

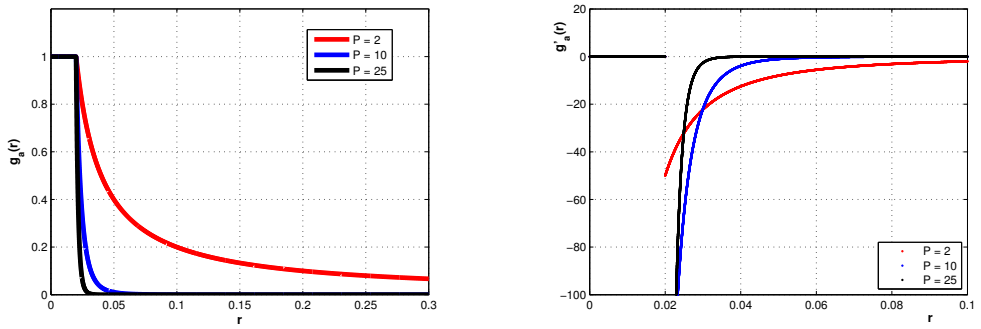
$$g_a(r) = \begin{cases} 1 & \text{if } 0 \leq r < \gamma \\ \left(\frac{\gamma}{r}\right)^{p/2} & \text{if } \gamma \leq r, \end{cases} \tag{16}$$

where $p > 1$ is the diffusivity parameter in (14), and γ is a threshold parameter that allows to get linear diffusion when r is less than γ, and backward diffusion when r is bigger than γ. Let us observe that μ_0 is related to γ as $\sqrt{\gamma} = \mu_0$.

In this way this diffusivity looks like the bounded step function $B(s)$ when $\mu_0 = \sqrt{\gamma}$ and $p \to \infty$, as it can be seen in Fig. 2(a). It can also be seen that g_a is continuous and differentiable at any point in $[0, +\infty)$, except at $r = \gamma$. But its derivative is bounded by $-\frac{p}{2}\gamma^{-1}$, so we can extend it to $r = \gamma$ by defining $g_a'(\gamma) \equiv \lim_{x \to \gamma^+} g_a'(x)$. Hence, it takes the form

$$\widehat{g_a'}(r) = \begin{cases} 0 & \text{if } r < \gamma \\ -\frac{p}{2}\,\frac{\gamma^{p/2}}{r^{p/2+1}} & \text{if } \gamma \leq r, \end{cases} \tag{17}$$

and therefore $g_a'(r) \overset{\text{a.e.}}{=} \widehat{g_a'}(r)$ (*a.e.*: almost everywhere) (see Fig. 2(b)). This extension is not particulary restrictive, from now on we will work with the function $\widehat{g_a'}$ as the derivative of g_a and, accordingly, we will use the notation g_a' instead of $\widehat{g_a'}$.



(a) g_a (b) g_a'

Figure 2: Graphics of the proposed diffusivity for different values of p (*e.g.* $\gamma = 0.02$).

We also note that g_a has the Lipschitz property.

Proposition 2. *The function g_a given by* (16) *is Lipschitz continuous in* $[0, +\infty)$ *(i.e.* $g_a \in \mathcal{C}^{0,1}([0, +\infty))$*)*

Proof. We have to show that

$$|g_a(r_2) - g_a(r_1)| \leq L\,|r_2 - r_1| \quad \forall r_1, r_2 \in [0, +\infty) \tag{18}$$

with $r_1 < r_2$. The proof will be divided into three steps: *(a)* $r_1, r_2 \in [0, \gamma]$, *(b)* $r_1, r_2 \in [\gamma, +\infty)$ and *(c)* $r_1 < r_2$ with $r_1 \in [0, \gamma)$ and $r_2 \in (\gamma, \infty)$. The proofs of the first two steps, are immediate using the Mean Value Theorem. We now analyze the third case. Let us assume that $r_1 < r_2$ with $r_1 \in [0, \gamma)$ and $r_2 \in (\gamma, \infty)$. According to the results for *(a)*, *(b)* and using that $|r_2 - \gamma| \leq |r_2 - r_1|$ we have that

$$\begin{aligned} |g_a(r_2) - g_a(r_1)| &= |g_a(r_2) - g_a(\gamma) + g_a(\gamma) - g_a(r_1)| \\ &\leq |g_a(r_2) - g_a(\gamma)| + |g_a(\gamma) - g_a(r_1)| \\ &\leq L\,|r_2 - \gamma| \leq L\,|r_2 - r_1| \end{aligned}$$

It may be concluded that $\forall\, r_1, r_2 \in [0, \infty)$, with $r_1 < r_2$, equation (18) is satisfied with $L = \frac{p}{2}\,\gamma^{-1}$.

□ □

From now on, to shorten and clarify the notation, we will write g instead of g_a for the function in (16).

4 Numerical approximation for 1D

4.1 Semi-discretization procedure for 1D

We begin using the MOL (Method of Lines) [22] for the equation (3) in 1D with the diffusivity g given by (16). Therefore, we follow a spatial semi-discretization procedure to convert a partial differential equation into a system of coupled ordinary differential equations. For simplicity, we will use a uniform grid, with a grid spacing h. The function $\boldsymbol{U}(t)$ will be obtained as the solution of a system of ordinary differential equations that comes from approximating $\partial_x u(x,t)$ at points halfway between the grid points, using a centered approximation in the original equation (3). In matrix form, the system of autonomous ordinary differential equations with Neumann boundary conditions is

$$\frac{d\boldsymbol{U}}{dt}(t) = \boldsymbol{f}(\boldsymbol{U}(t)) = \boldsymbol{A}(\boldsymbol{U}(t))\,\boldsymbol{U}(t) \tag{19}$$

where $\boldsymbol{A} = [a_{ij}]$ has the following coefficients, for $i = 2, \ldots, m-1$

$$a_{i,i+1} = \frac{g_i}{h^2} = a_{i+1,i}; \quad a_{i,i} = -\frac{g_{i-1} + g_i}{h^2} \tag{20a}$$

$$a_{i,i-1} = \frac{g_{i-1}}{h^2} = a_{i-1,i} \tag{20b}$$

being $g_i = g\left(\frac{1}{h^2}|U_{i+1} - U_i|^2\right)$ with $i = 1, \ldots, m-1$ and for the Neumann boundary conditions

$$a_{1,1} = -\frac{1}{h^2}\, g_1; \quad a_{m,m} = -\frac{1}{h^2}\, g_{m-1}$$

and $a_{i,j} = 0$ otherwise.

The next step is to check whether the system of differential equations (19), with the diffusivity function proposed in (16), satisfies the well-posedness and semi-discrete scale-space requirements [43]. For this reason, the matrix $\boldsymbol{A}(\boldsymbol{U})$ must meet certain requirements as were established in [43] by Weickert (see $(S1)$ to $(S5)$ in page 76). It can be easily proven that $\boldsymbol{A}(\boldsymbol{U})$ is a real tridiagonal and symmetric matrix ([43, $(S2)$]). As a consequence of the positiveness of diffusitivity ($g > 0$) given by (16), $a_{i,\,i+1}$ and $a_{i+1,\,i}$ are different from zero and non-negative ([43, $(S4)$]). Hence the matrix is irreducible [38] ([43, $(S5)$]). It is clear that the rows sum are zero and the off-diagonal terms are non-negative ([43, $(S3)$]). Now we show that $\boldsymbol{A}(\boldsymbol{U})$ has the Lipschitz condition for (16) for every bounded subset ([43, $(S1)$]).

Proposition 3. *The matrix* $\boldsymbol{A}(\boldsymbol{U}) = [a_{ij}(\boldsymbol{U})]$ *defined in* (19) *with the diffusivity function* g (16)*, satisfies the Lipschitz condition for any bounded subset of* $\mathbb{R}^m$*.*

Proof. For any $K > 0$ we consider the set given by

$$\Upsilon = \{\boldsymbol{U} \in \mathbb{R}^m \mid \|\boldsymbol{U}\|_1 \leq K\} \tag{21}$$

By the Proposition 2 above,

$$\left| g\left(\frac{1}{h^2}(U_{i+1} - U_i)^2\right) - g\left(\frac{1}{h^2}(V_{i+1} - V_i)^2\right)\right| \leq \left|\frac{L}{h^2}\left[(U_{i+1} - U_i)^2 - (V_{i+1} - V_i)^2\right]\right|$$

but

$$\begin{aligned}\left[(U_{i+1} - U_i)^2 - (V_{i+1} - V_i)^2\right] &= [(U_{i+1} - U_i) + (V_{i+1} - V_i)] \\ &\quad \cdot [(U_{i+1} - V_{i+1}) + (V_i - U_i)]\end{aligned}$$

and

$$0 \leq |U_{i+1} - U_i| \leq K \quad i = 1, \ldots, m-1 \quad \forall\, \boldsymbol{U} \in \Upsilon$$

hence

$$\left| g\left(\frac{1}{h^2}(U_{i+1} - U_i)^2\right) - g\left(\frac{1}{h^2}(V_{i+1} - V_i)^2\right)\right| \leq \frac{2K\,L}{h^2}\Big[\,|U_{i+1} - V_{i+1}| + |U_i - V_i|\,\Big]. \tag{22}$$

Now, let us denote by $\boldsymbol{A}(\boldsymbol{U}) = [a_{i,j}]$ and $\boldsymbol{A}(\boldsymbol{V}) = [\tilde{a}_{i,j}]$, then

$$\|\boldsymbol{A}(\boldsymbol{U}) - \boldsymbol{A}(\boldsymbol{V})\|_1 = \max_j \sum_{i=1}^{m} |a_{i,j} - \tilde{a}_{i,j}|.$$

The two matrices are symmetric and tridiagonal, therefore

$$\begin{aligned}\|\boldsymbol{A}(\boldsymbol{U}) - \boldsymbol{A}(\boldsymbol{V})\|_1 = \max_j \big[&|a_{j,j-1} - \tilde{a}_{j,j-1}| + |a_{j,j} - \tilde{a}_{j,j}| \\ &+ |a_{j,j+1} - \tilde{a}_{j,j+1}|\big]\,.\end{aligned}$$

According to (20) and (22), we get:

$$|a_{j,j} - \tilde{a}_{j,j}| \leq \frac{1}{h^2}\,|g_{j-1} - \tilde{g}_{j-1}| + \frac{1}{h^2}\,|g_j - \tilde{g}_j|$$
$$\leq \frac{2\,K\,L}{h^4}\Big[\,|U_{j+1} - V_{j+1}| + 2\,|U_j - V_j| + |U_{j-1} - V_{j-1}|\,\Big]$$
$$\leq \frac{4\,K\,L}{h^4}\|\boldsymbol{U} - \boldsymbol{V}\|_1.$$

Similar arguments apply to the cases

$$|a_{j,j-1} - \tilde{a}_{j,j-1}| \leq \frac{2\,L}{h^4}\|\boldsymbol{U} - \boldsymbol{V}\|_1 \quad \text{and}$$
$$|a_{j,j+1} - \tilde{a}_{j,j+1}| \leq \frac{2\,K\,L}{h^4}\|\boldsymbol{U} - \boldsymbol{V}\|_1.$$

Finally, we get

$$\|\boldsymbol{A}(\boldsymbol{U}) - \boldsymbol{A}(\boldsymbol{V})\|_1 \leq\ M\|\boldsymbol{U} - \boldsymbol{V}\|_1$$

being

$$M = \frac{8\,K\,L}{h^4} \quad \text{with} \quad L = \frac{p}{2}\,\gamma^{-1}$$

Since there has been no restriction on the set Υ, it must be true for any bounded set of $\mathbb{R}^m$.

□ □

Therefore, following the results by Weickert and Benhamouda in [45], the semidiscrete (19) process for diffusivities g (16), satisfies the well-posedness and semi-discrete scale-space requirements [43].

4.2 Full discretization for 1D

We now proceed to discretize the temporal variable in the ordinary differential system (19). We have two possibilities, one is to set an explicit scheme. The drawback of this method is the restriction for time step size [46]. For this reason, we prefer to use implicit schemes. In particular we use the Backward Euler method. In this case we have:

$$\frac{\boldsymbol{U}^{n+1} - \boldsymbol{U}^n}{k} = \boldsymbol{A}(\boldsymbol{U}^{n+1})\boldsymbol{U}^{n+1} = \boldsymbol{f}(\boldsymbol{U}^{n+1}) \tag{23}$$

where $\boldsymbol{f}$ is given by (19) and we approximate $\boldsymbol{U}(t_n) = \boldsymbol{U}(nk)$ by $\boldsymbol{U}^n$ with $k \equiv \Delta t$, the time step size. We can easily realize that (23) is a nonlinear algebraic equation. One straightforward possibility to solve it, is to use the Newton-Raphson method. In this case, we consider the function

$$\boldsymbol{q}(\boldsymbol{U}) = \boldsymbol{U} - k\,\boldsymbol{f}(\boldsymbol{U}) - \boldsymbol{U}^n = [\boldsymbol{I} - k\,\boldsymbol{A}(\boldsymbol{U})]\,\boldsymbol{U} - \boldsymbol{U}^n \tag{24}$$

and the Newton-Raphson's method is the iterative method

$$\boldsymbol{U}^{\nu+1} = \boldsymbol{U}^{\nu} - \left(\boldsymbol{I} - k\,\frac{\partial\,\boldsymbol{f}}{\partial \boldsymbol{U}}(\boldsymbol{U}^{\nu})\right)^{-1} \boldsymbol{q}(\boldsymbol{U}^{\nu}) \tag{25}$$

with $\nu = 0, 1, \ldots$ and where $\left(\boldsymbol{I} - k\,\dfrac{\partial\,\boldsymbol{f}}{\partial \boldsymbol{U}}(\boldsymbol{U}^{\nu})\right)$ is the iteration matrix. To construct explicitly this iteration matrix we give an expression for the Jacobian matrix of $\boldsymbol{f}$ given by (19).

Proposition 4. *The Jacobian matrix of $\boldsymbol{f}$ defined by* (19) *can be expressed by a sum of two real tridiagonal and symmetric matrices*

$$\frac{\partial\,\boldsymbol{f}}{\partial \boldsymbol{U}}(\boldsymbol{U}) = \boldsymbol{A}(\boldsymbol{U}) + \boldsymbol{C}(\boldsymbol{U}). \tag{26}$$

Proof. By definition of function the $\boldsymbol{f}(\boldsymbol{U}) = \boldsymbol{A}(\boldsymbol{U})\,\boldsymbol{U}$ and using derivation rules, yields

$$\boldsymbol{\mathfrak{J}} = \frac{\partial\,\boldsymbol{f}}{\partial \boldsymbol{U}}(\boldsymbol{U}) = \boldsymbol{A}(\boldsymbol{U}) + \frac{\partial \boldsymbol{A}(\boldsymbol{U})}{\partial \boldsymbol{U}}\boldsymbol{U}.$$

For simplicity of notation, we use[2]

$$\boldsymbol{C}(\boldsymbol{U}) \equiv \frac{\partial \boldsymbol{A}(\boldsymbol{U})}{\partial \boldsymbol{U}}\,\boldsymbol{U}, \quad \text{so} \quad [c_{i,l}] = \left[\frac{\partial a_{i,j}}{\partial U_l}\,U_j\right],$$

and $f_{i,j}$ for the Jacobian coefficients. It is easy to check that

$$\frac{\partial a_{i,i-1}}{\partial U_l} = 0 \quad \text{if } l \neq i-1 \text{ or } l \neq i,$$

since $a_{i,i-1} = (1/h^2)\,g_{i-1}$ only depends on U_i and U_{i-1}. Similarly

$$\frac{\partial a_{i,i+1}}{\partial U_l} = 0 \quad \text{if } l \neq i+1 \text{ or } l \neq i.$$

[2]Note Einsten's criterion

This gives the follow terms, not necessarily null:

$$\begin{aligned} c_{i,i+1} &= \frac{2}{h^4}\, g_i' \cdot (U_{i+1} - U_i)^2 \quad \text{with } i = 1, \ldots, m-1, \\ c_{i,i} &= -\frac{2}{h^4}\left[g_i' \cdot (U_{i+1} - U_i)^2 + g_{i-1}' \cdot (U_i - U_{i-1})^2 \right] \\ & \quad \text{with } i = 2, \ldots, m-1, \\ c_{i,i-1} &= \frac{2}{h^4}\, g_{i-1}' \cdot (U_i - U_{i-1})^2 \quad \text{with } i = 2, \ldots, m \end{aligned}$$

and

$$\begin{aligned} c_{1,1} &= -\frac{2}{h^4}\, g_1' \cdot (U_2 - U_1)^2; \\ c_{m,m} &= -\frac{2}{h^4}\, g_{m-1}' \cdot (U_m - U_{m-1})^2, \end{aligned}$$

where

$$g_i' = g'\left(\frac{1}{h^2} (U_{i+1} - U_i)^2 \right) \qquad i = 1, \ldots, m-1. \tag{27}$$

We also note that $c_{i,i+1} = c_{i+1,i}$ since

$$\begin{aligned} \frac{\partial\, a_{i,i+1}}{\partial\, U_{i+1}} U_{i+1} = \frac{\partial\, a_{i+1,i+1}}{\partial\, U_i} U_{i+1}, \quad \text{and} \\ \frac{\partial\, a_{i,i}}{\partial\, U_{i+1}} U_i = \frac{\partial\, a_{i+1,i}}{\partial\, U_i} U_i. \end{aligned}$$

Similar arguments apply to the case $c_{i-1,i} = c_{i,i-1}$. Consequently we can establish the relation (26).

□ □

It is worth pointing out several consequences that can be extracted from the last proposition. Firstly, the expression (26) is directly connected with the term

$$g(\|\nabla u\|^2) + 2\, g'(\|\nabla u\|^2) \|\nabla u\|^2 \tag{28}$$

appearing in equations (5) or (8b). That is, according to the matrix elements $a_{i,j}$ and $c_{i,j}$ of the expression (26) we can relate these terms with (28) in the following: the coefficients of matrix (A) only use the function g, as we can see in (20), while the coefficients of matrix (C) only use the function g', as we can see in the proof of Proposition 4. Secondly, the matrix $\boldsymbol{C}(\boldsymbol{U})$ is well

defined because the derivative of the proposed diffusivity (16) exists at any point in $[0, \infty)$ as we have established in (17).

On the other hand, we also note that if g is a monotone decreasing function, its derivative g' is non-positive. This implies that the $c_{i,j}$ coefficients are also non-positive and the coefficients of the Jacobian could change their signs and thus the signs of the eigenvalues of the Jacobian matrix in the iterative process. This can make the iteration matrix singular and, hence, some important instabilities in the numerical solution might appear.

To avoid this drawback, we apply the *tangential stiffness method* that will connect with an edge-preserving process. This scheme has already been used in nonlinear Finite Element analysis context for nonlinear mechanics of solids and plasticity problems [28, 31]. We propose using this method to solve the nonlinear equation (23). We give the following approximation to the Newton-Raphson iteration matrix (25)

$$\boldsymbol{I} - k\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{U}}(\boldsymbol{U}) = \boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U}) - k\boldsymbol{C}(\boldsymbol{U}) \approx \boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U}) \tag{29}$$

Substituting this approximation in expression (25) and using (24), we arrive at the following iterative process:

$$\boldsymbol{U}^{\nu+1} = \left[\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U}^{\nu})\right]^{-1}\boldsymbol{U}^{n} \tag{30}$$

with $\nu = 0, 1, 2, \ldots \quad \forall\, k \in \mathbb{R}^{+}$ where we choose $\boldsymbol{U}^{n}$ for $\nu = 0$ as first iteration term. The advantage of using this iteration matrix lies in the fact that $\boldsymbol{W}_k(\boldsymbol{U}) \equiv [\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U})]$ is positive definite by the properties of $g > 0$ for all $\boldsymbol{U}$ and $k \in \mathbb{R}^{+}$ (see Ortega [27, p. 107]).

Since $\boldsymbol{W}_k(\boldsymbol{U})$ is positive definite, $\boldsymbol{W}_k(\boldsymbol{U})^{-1}$ exists. We also know that $\boldsymbol{W}_k(\boldsymbol{U})$ satisfies $w_{i,j} \leq 0$, $i \neq j$ and $w_{i,i} > 0$, $i = 1, \ldots, m$ and is symmetric, then by [27, p. 110], it is a *Stieltjes matrix* so, $\boldsymbol{W}_k(\boldsymbol{U})^{-1} \geq O$ (see [27]).

We now consider $[\boldsymbol{V}]^t = [1, 1, \ldots, 1]$ and since $\boldsymbol{A}(\boldsymbol{U})$ has unit row sum, we have $[\boldsymbol{V}]^t\boldsymbol{W}_k(\boldsymbol{U}) = [\boldsymbol{V}]^t$ and it yields $[\boldsymbol{V}]^t\boldsymbol{W}_k(\boldsymbol{U})^{-1} = [\boldsymbol{V}]^t$. Then we also can conclude that

$$\|\boldsymbol{W}_k(\boldsymbol{U})^{-1}\|_1 \equiv \left\| \left[\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U})\right]^{-1} \right\|_1 = 1 \quad \forall\, k \in \mathbb{R}^{+}$$

Because of $\left[\boldsymbol{I} - k\,\boldsymbol{A}\,(\boldsymbol{U})\right]^{-1}$ is symmetric, we also have $\left\|\left[\boldsymbol{I} - k\,\boldsymbol{A}\,(\boldsymbol{U})\right]^{-1}\right\|_{\infty} = 1$

Proposition 5. *For the iterative process* (30)*, we have*

$$\|\boldsymbol{U}^{\nu+1}\|_1 \leq \|\boldsymbol{U}^n\|_1 \qquad \nu = 0, 1, 2, \ldots \qquad \forall\, k \in \mathbb{R}^+$$

where we choose $\boldsymbol{U}^n$ *as the first iteration.*

Proof. The proof is straightforward. By

$$\left\|[\boldsymbol{I} - k\,\boldsymbol{A}\,(\boldsymbol{U})]^{-1}\right\|_1 = 1 \quad \text{it follows that}$$

$$\|\boldsymbol{U}^{\nu+1}\|_1 \leq \left\|[\boldsymbol{I} - k\,\boldsymbol{A}\,(\boldsymbol{U}^{\nu})]^{-1}\right\|_1 \|\boldsymbol{U}^n\|_1 = \|\boldsymbol{U}^n\|_1$$

with $\nu = 0, 1, 2, \ldots \quad \forall\, k \in \mathbb{R}^+$. This establishes that the iterative process is stable $\forall\, k \in \mathbb{R}^+$.

□ □

The iterative scheme (30) is also known as the *direct iteration method*, the *successive approximation method* or the *Picard method* [28, 31].

We emphasize several points about the iterative process in (30).

1. If we only take one iteration, we get the *semi-implicit method* [46]. In this case, from the above results, it can directly prove that the iterative process in (30), with g given by (16), verifies the following criterion [46]: $(D1)$ *Continuity in its argument*, $(D2)$ *Symmetry*, $(D3)$ *Unit row sum*, $(D4)$ *Positive diagonal*, $(D5)$ *Irreducibility*. Under these conditions, the iterative process (30) with only one time step, creates a discrete scale-space [43]. So, it has proved that the function of diffusivity proposed in (16) satisfies the well-posedness and semi-discrete and full discrete scale-space requirements.

2. According to the approximation given by (29) we do not construct the inverse diffusion numerically, we only do forward diffusion. We do not use the matrix $\boldsymbol{C}$ because it involves the derivative of the diffusivity g and it is non-positive as we have mentioned before. However, g is a decreasing function, which implies diffusion with different speed. That is, for those pixels that satisfy the condition $\|\nabla u\|^2 < \gamma$, the proposed diffusion function g (13) reaches its maximum value, $g = 1$, we get the linear diffusion. However, when we have pixels that satisfy the condition $\gamma \leq \|\nabla u\|^2$, then the proposed diffusion function g (13) takes a value between 0 and 1 and it produces slower diffusion for those pixels. Thus, two contiguous pixels can be separated because one of them can produce a value of $g = 1$ and the other one $0 < g < 1$. This is a consequence of using an edge-preserving process.

3. From the iterative process (30), we can establish the following iterative scheme in the compact and convex set Υ (21)

$$\boldsymbol{U}^{\nu+1} = \boldsymbol{G}(\boldsymbol{U}^{\nu}), \text{ with } \boldsymbol{G}(\boldsymbol{U}^{\nu}) = [\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U}^{\nu})]^{-1}\boldsymbol{U}^{n}, \qquad (31)$$

for $\nu = 0, 1, 2 \ldots$ that easily yields a *fixed point iteration*. Obviously a fixed point for (31), that is $\boldsymbol{U}^* = \boldsymbol{G}(\boldsymbol{U}^*)$ is equivalent to the equation $\boldsymbol{q}(\boldsymbol{U}^*) = \boldsymbol{0}$. It is not our purpose to study this iterative process when $\nu \geq 1$, because it is not clear that this discrete scheme could be connected with a filtering process and create a discrete space-scale [43, 46]. However, this result is interesting because it allows us to say that the iterative process (31) has at least one solution, without putting any restrictions on the parameter k.

5 Numerical approximation for N dimension

We now extend the iterative process (30) for two and three space dimensions and in the remainder of this section we assume g to be the diffusivity defined by (16). In this case, we follow a similar process as in the case of 1D.

5.1 Semi-discretization

We again apply the *method of lines* to the equation (3) for 2-D or 3-D. On the rectangular or brick domain we establish a spatial grid. Therefore, the semidiscrete system has dimension $m = m_1 \cdots m_s$, where $s = 2$ or 3. Again, solving the semi-dicrete problem means finding a function $\boldsymbol{U}(t)$ which is approximation to the solution u at discrete grid points $(x_1, \ldots, x_s, t)$. The function $\boldsymbol{U}(t)$ will be obtained as the solution of a system of ordinary differential equations that it comes from approximating $\partial_{x_r} u_\alpha$ (where $r = 1, \ldots, s$ and $\alpha = (i, j)$ or $\alpha = (i, j, k)$) at points halfway between the grid points, using a centered approximation in the original equation (3).

We only use one index for pixel numbering, so we can represent the whole image of size $m_1 \times \cdots \times m_s$ by a vector of size $m = m_1 \cdots m_s$. In this case, we have again an ordinary differential system of type (19) where the coefficient matrix $\boldsymbol{A}(\boldsymbol{U})$ is now a block tridiagonal matrix of order m. It is straightforward to check that $\boldsymbol{A}(\boldsymbol{U})$ satisfies the conditions of Lipschitz-continuity, symmetry, vanishing row sums, nonnegative off-diagonals and finally, because of its associated directed graph is strongly connected (see Ortega [27]),

it verifies the condition of irreducibility. Therefore, this differential system of type (19) also satisfies the prerequisites to be a filtering process (see Weickert [43]).

It is worth pointing out that $\boldsymbol{A}(\boldsymbol{U})$ can be decomposed in the form

$$\boldsymbol{A}(\boldsymbol{U}) = \sum_{r=1}^{s} \boldsymbol{A}_r(\boldsymbol{U}) \tag{32}$$

where each $\boldsymbol{A}_r(\boldsymbol{U})$ represents one-dimensional semi-discretization of the diffusion process along the x_r axes. It is immediate that $\boldsymbol{A}_r(\boldsymbol{U})$ with $r = 1, \ldots, s$ verify again the properties of Lipschitz-continuity, symmetry, vanishing row sums and nonnegative off-diagonals.

5.2 Time discretization

We can extend to 2D or 3D the full discretization process developed for 1D. Therefore, from the semi-discrete system (19) with the block-tridiagonal matrix $\boldsymbol{A}(\boldsymbol{U})$ as coefficient matrix, we arrive to the nonlinear equation (23). We also use the *tangential stiffness method* to solve it. It yields to a similar iterative process (30) that we have established for 1D:

$$\boldsymbol{U}^{\nu+1} = \Big[\boldsymbol{I} - k\,\boldsymbol{A}(\boldsymbol{U}^{\nu})\Big]^{-1}\boldsymbol{U}^{n} \quad \text{with} \;\; \nu = 0, 1, \ldots \tag{33}$$

It is straightforward to show that the *iteration matrix* $\boldsymbol{W}_k(\boldsymbol{U}) = \boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U})$ has also the properties that we have proved for 1D and it is immediate to check that $\| \left[\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U})\right]^{-1} \|_1 = 1$ too. Finally, it follows easily that $\left[\boldsymbol{I} - k\boldsymbol{A}(\boldsymbol{U}^{\nu})\right]^{-1}$ satisfies the requirements of continuity in its arguments, symmetry, unit row sum, positive diagonal and irreducibility that it can be done in a similar way as the 1D case. Thus, for only one iteration, that is: $\boldsymbol{U}^n$ for $\nu = 0$ (semi-implicit method) we get the scale-space requirements [43].

One possibility to solve the iterative process (33) is the iterative scheme

$$\boldsymbol{U}^{\nu+1} = \frac{1}{s}\sum_{r=1}^{s} \left[\boldsymbol{I} - s\,k\boldsymbol{A}_r\left(\boldsymbol{U}^{\nu}\right)\right]^{-1}\left[\boldsymbol{U}^{n}\right] \equiv \boldsymbol{Q}\left(\boldsymbol{U}^{\nu}\right)\left[\boldsymbol{U}^{n}\right] \tag{34}$$

with $\nu = 0, 1, \ldots$ where we get the *AOS scheme* [46] for $\nu = 0$.

For one iteration, we take $\boldsymbol{U}^n$ for $\nu = 0$, then $\boldsymbol{Q}(\boldsymbol{U}^n)$ satisfies again the properties of continuity in its arguments, symmetry, unit row sum, positive

diagonal and irreducibility and consequently the iterative process $\boldsymbol{U}^{n+1} = \boldsymbol{Q}(\boldsymbol{U}^n)[\boldsymbol{U}^n]$ creates a discrete scale-space, for the propose diffusivity g (16).

It is worth pointing out that the same conclusion that we got for 1D can be drawn for 2D and 3D. That is, as a consequence of the iterative process (33), we again do not use the derivative of g for discrete scheme because we do not use the matrix $\boldsymbol{C}$, so we do not backward diffusion from the numerical point of view, but forward diffusion with a decreasing filter g.

On the other hand, with the discrete scheme (34), we only can ensure again, a discrete scale-space process for $\nu = 0$ and for each time step $k > 0$, as we did for 1D.

6 Experiments

In this section, we report some experimental results obtained when we apply our filter to natural gray-scale images chosen from a public database [13]. The aim of the experiments is to show the performance of our filter to obtain piecewise constant images as a previous step to the segmentation process. In the proposed method, four key parameters are investigated, the first two are specific to the numerical method: the stopping time T and the time step k. The other two are needed to define the diffusivity: the edge strength threshold γ and the parameter p.

We take into account a two dimensional discrete image of size $m_1 \times m_2$ as a vector $\boldsymbol{U} \in \mathbb{R}^m$, $m = m_1 \times m_2$, whose components, U_i, $i = 1, \ldots, m$, represent the grey values at each pixel. Let $\boldsymbol{U}^{(0)}$ be the original image, then the discrete scale-space given by (34), is used to calculate a sequence $(\boldsymbol{U}^{(j)})_{j\in\mathbb{N}}$ of filtered versions of $\boldsymbol{U}^{(0)}$ by gradually removing noise and details until the image converges to the average grey level $\mu = \frac{1}{m}\sum_{i=1}^{m} U_i^{(0)}$, as $t \to +\infty$.

6.1 Stopping time T and time step k

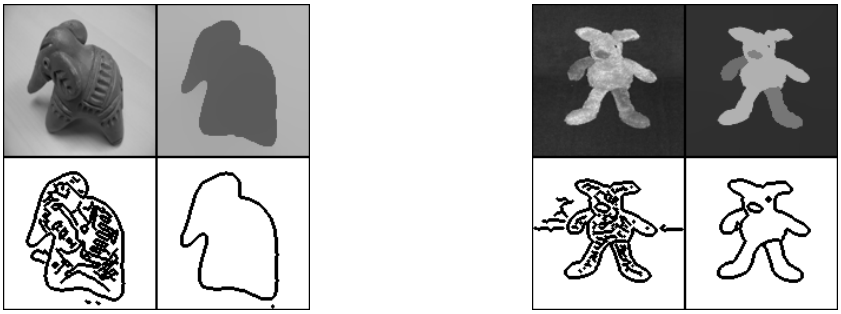
For nonlinear diffusion, the stopping time has a strong effect on the diffusion result and, as far as we know, there is not a straightforward stopping criterion. Several studies have addressed the stopping time problem in the field of PDEs image restoration. In [14–17, 19, 25, 30, 32, 34, 37, 44], criterion on the stopping time selection are closely linked to the noise-filtering problem. In these works, the aim is to stop the process before the structure of the image has been modified too much while textured structures must be maintained.

In our case, the aim is to achieve piecewise constant images as a previous process of segmentation techniques, therefore we must determine a stopping time not only to eliminate image noise but noise and texture as well. The proposed stopping time takes into account that the diffusivity we use is *linear* when the modulus of the gradient is less than $\sqrt{\gamma}$, and tends to zero when the modulus of the gradient is greater than $\sqrt{\gamma}$, thus instead of solving the nonlinear problem, we solve the linear diffusion problem and consider the setting time, t_s,

$$\frac{\|\boldsymbol{U}^{(n)} - \boldsymbol{\mu}\|}{\|\boldsymbol{\mu}\|} \leq 0.02 \qquad t_s = n \cdot k \tag{35}$$

where $\boldsymbol{\mu} = (\mu, \ldots, \mu)^T \in \mathbb{R}^m$, n is the number of iterations performed and $\boldsymbol{U}^{(n)}$ is the filtered image using linear diffusion. We note that the proposed stopping time, $T = t_s = n \cdot k$, operates without prior knowledge of the geometric or statistical structure of the image because it depends only on the size and pixel intensities of the initial image $\boldsymbol{U}^{(0)}$. Since the amount of work involved in the method is proportional to the number of individual steps, we attempt to choose the time step k as large as possible, but such that no numerical artifacts appear when we solve the nonlinear problem using the decomposition method AOS. Experimentally we found that the time step size can take high values without noticeable difference in the results.

Given the proposed nonlinear diffusion filter, an original image is composed of several regions and choose the parameters p and γ as described in the following subsection, the filtered image obtained on the scale t_s is an image in which regions are homogenized, while the edges are not blurred by the strength of the diffusion function. For example, the regions of the images in Fig. 3 are homogenized with the proposed filter while the edges are not blurred.



(a) (b)

Figure 3: Top row: The left one, show the original image: $\boldsymbol{U}^{(0)}$. The right one, show the last iteration performed: $\boldsymbol{U}^{(n)}$. The image size was 128×128 and the number of iterations, n, was obtained from (35). Bottom row: The Canny detector, ($\sigma = 1$), in order to compare the edges detected in the initial image and the filtered image.

6.2 Diffusivity parameters: γ and p

The parameter γ has been chosen taking into account the statistical interpretation of nonlinear diffusion from the point of view of robust statistics developed in [5]. The authors show that nonlinear diffusion can be seen as a robust estimation procedure that estimates a piecewise constant image from a noisy image where the boundaries between the piecewise regions are considered to be outliers. We apply this statistical interpretation to select $\gamma = (1.4826 \cdot MAD(\|\nabla(\boldsymbol{U}^{(0)})\|))^2$, where MAD is the median absolute deviation.

The diffusivity is defined so that the filter is linear when $\|\nabla(\boldsymbol{U}^{(j)})\| \leq \sqrt{\gamma}$ and blurs the regions and when $\|\nabla(\boldsymbol{U}^{(j)})\| > \sqrt{\gamma}$, the diffusion function $g(\|\nabla(\boldsymbol{U}^{(j)})\|)$ tends to zero as shown in Fig. 2(a). We also note that high p values in (16) prevent edge blurring because p tends to the *bounded step function* (Fig. 1). Then, in the experiments, a piecewise constant image should be obtained for low values of p if the contrast among the objects and the background is high. The value of p should be higher if the image has more details or the contrast between the objects and the background is low. We apply the diffusion model to three images in order to show the effects of different values of p.

In Fig. 4 we see that the background of the flower image has more details than the bear image, then for the bear we obtained piecewise constant images when $p = 2.5$ or 3 and for the flower when $p = 5$ or 5.5. In the case of the bush image, Fig. 5 shows that the contrast among the different objects is low and to avoid blurring the value of p must be greater than in previous cases, $p = 19$ or 20.

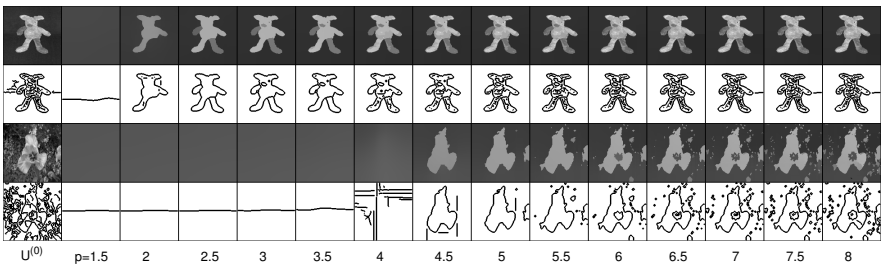


Figure 4: For $p = 1.5, 2, 2.5, \ldots, 8$ we show the original images $\boldsymbol{U}^{(0)}$ and $\boldsymbol{U}^{(n)}$ which correspond to the last iteration performed. The image size was 128×128 and the number of iterations, n, was obtained from (35). The time step was $k = 200$ and the iterations performed were $n = 39$ for the bear image and $n = 42$ for the flower image. The edges were detected using Canny detector ($\sigma = 1$).

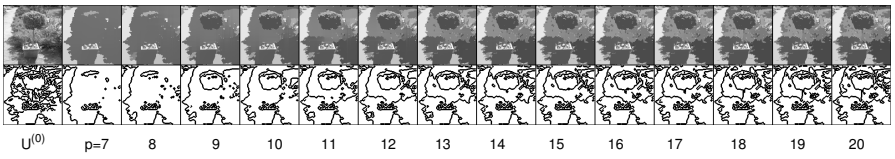


Figure 5: For $p = 7, 8, \ldots, 20$ we show the original bush image $\boldsymbol{U}^{(0)}$ and $\boldsymbol{U}^{(n)}$ which corresponds to the last iteration performed. The image size was 128×128 and the number of iterations, n, was obtained from (35). The time step was $k = 200$ and the iterations performed were $n = 51$. The edges were detected using Canny detector ($\sigma = 1$).

6.2.1 Proposal tuning parameter p by using prior knowledge

To select the value of p, we have used an edge detector and a training image database with their manual segmentation, as ground truth. Fig. 6 shows some manual segmentations used for tuning the value of p. Then we have

considered the *precision* and *recall* measures to calculate the *F-measure* [23, 36]. Let E_I be the set of edge pixels for the manual segmentation and E_F^p the set of edge pixels of the filtered image,

$$P(p) = |E_I \cap E_F^p|/|E_F^p| \qquad R(p) = |E_I \cap E_F^p|/|E_I|$$

where $|\cdot|$ is the counting measure. That is, the *precision* $P(\cdot)$, is the ratio of the true edge pixels detected to the total number of edge pixels detected, and the *recall* $R(\cdot)$ is the ratio of the true edge pixels detected to the number of edge pixels of the manual segmentation.

The *F-measure* is defined as the weighted harmonic mean of precision and recall,

$$F(p, \alpha) = P(p)R(p)/(\alpha R(p) + (1 - \alpha)P(p)) \;\; \alpha \in [0, 1]$$

We chose: $p^* = \arg\max\{F(p, 1/2) : p > 1\}$. In the experiments, we have used $p = 1.5, 2, 2.5, \ldots, 20$ and $\alpha = 1/2$ to attach equal importance to precision and recall.

Figure 6: Ground truth images.

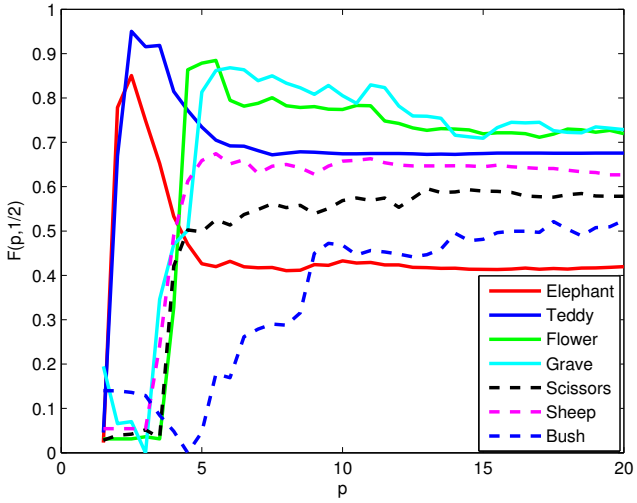


Figure 7: F-measure, $F(p, 1/2)$, is shown for $p = 1.5,\ 2,\ 2.5, \ldots, 20$.

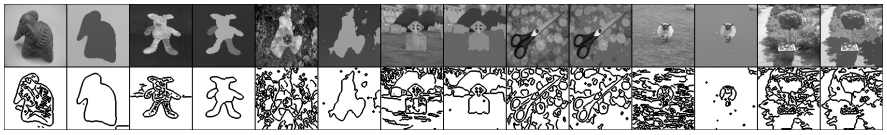


Figure 8: The original image $\boldsymbol{U}^{(0)}$ and the image corresponding to the last iteration $\boldsymbol{U}^{(n)}$ are shown here. The size of all images is 128×128 and $k = 200$. The p^* values and the iterations performed are: elephant $p^* = 2.5$, $n = 39$, bear $p^* = 2.5$, $n = 39$, flower $p^* = 5.5$, $n = 42$, grave $p^* = 6$, $n = 35$, scissors $p^* = 13$, $n = 41$, sheep $p^* = 5.5$, $n = 32$ and bush $p^* = 20$, $n = 51$. The edges were detected using Canny detector ($\sigma = 1$).

Table 1: The first row shows, F_0, the value of the F-measure for the original images. The second row shows $p^* = \arg\max\{F(p, 1/2) : p = 1.5, 2, \ldots, 20\}$. The third row shows the maximum value of the F-measure for p^*.

	Images						
	Bear	Bush	Elephant	Flower	Grave	Scissors	Sheep
F_0	0.597	0.578	0.379	0.573	0.606	0.704	0.623
p^*	2.5	20	2.5	5.5	6	13	5.5
$F(p^*, 1/2)$	0.918	0.554	0.852	0.913	0.832	0.693	0.699

Fig. 8 shows the performance of our filter with the proposed parameters. When it is applied to images with large and high contrast regions (elephant, bear, flower and grave) we see in Fig. 7 that the F-measure achieves an absolute maximum, edges are well preserved, no blurring is introduced and piecewise quasi-constant images are achieved. When the filter is applied to images in which the contrast between the object and the background is low or images with small objects and many local details along with edges that have low gradient (scissors, bush), the F-measure reaches a maximum value for a parameter p high, in order to avoid blurring of the edges, but if we compare the maximum value for these cases there is not a significant improvement with the value of F-measure for the initial image, see Tab. 1. In such cases in order to improve the outcomes the use of local γ has to be carefully analyzed.

Fig. 9 and Table 2 illustrate a comparative with two filter for obtaining piecewise constant images: (a) the cartoon variation image decomposition [3] and (b) a nonlinear diffusion filter which uses the AOS scheme [46]. These approaches are chosen because their tasks are the processing of piecewise constant images and these filters have very few parameters to adjust. The first one only requires a weighting between the fidelity term and TV regularization. In the second approach, the diffusivity function is replaced in the proposal and the numerical scheme is the same. Experimental results show: (a) the stopping time procedure is robust. The diffusivity function is changed in the numerical scheme and the piecewise constant images are obtained. (b) The proposed approach gives better qualitative results. High contrast details are kept unlike other two approaches and (c) the F-measures of our approach are higher than in the other two filters. The scripts used in this study are available at `https://github.com/cplatero/NonlinearDiffusion`.

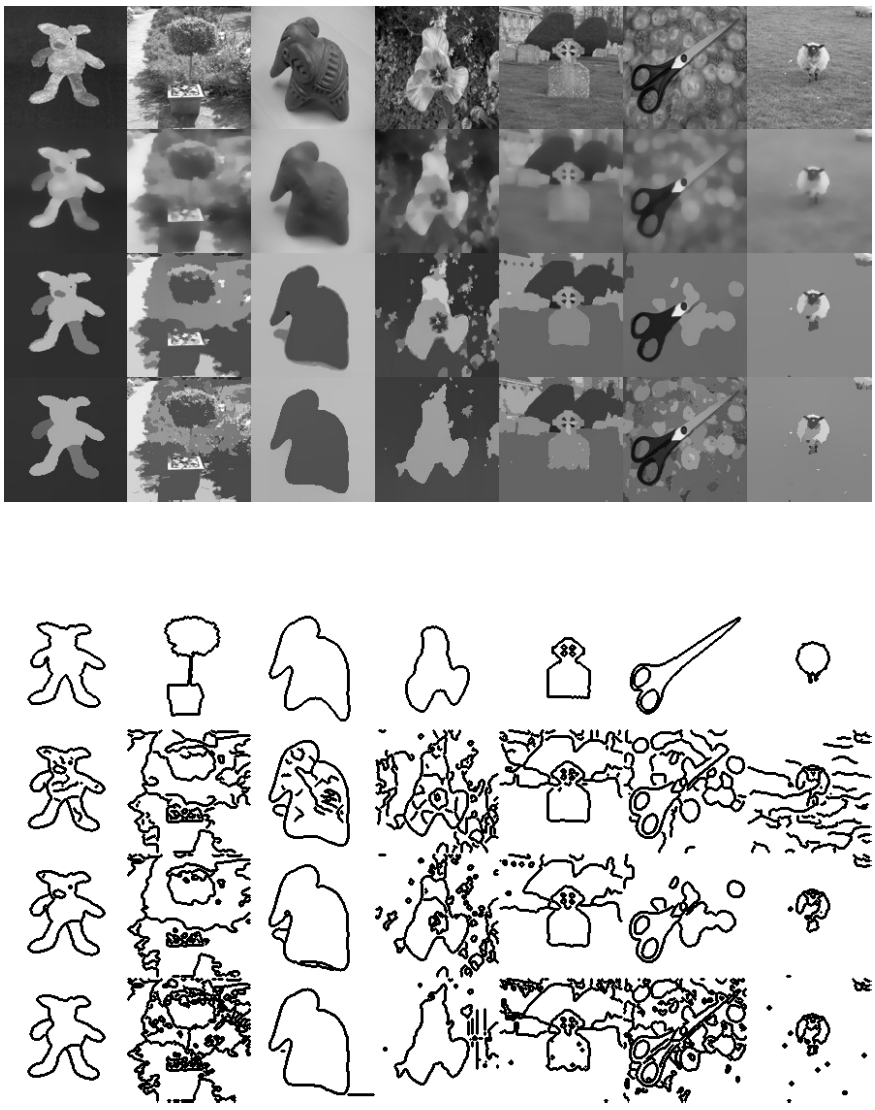


Figure 9: The first block shows the filtered images and the second block illustrates their edge detections. Within each block, the first row shows the original images. The next rows illustrate the output of the image filtering with the cartoon variation image decomposition, the nonlinear diffusion filter and our proposal, respectively.

Table 2: A quantitative comparative of the image filtering with three approaches are reported using the F-measures. The first row shows the F-measure in the original images. The next rows illustrate the the F-measure in the cartoon variation image decomposition, the nonlinear diffusion filter and our proposal, respectively.

Images							
	Bear	Bush	Elephant	Flower	Grave	Scissors	Sheep
F_0	0.597	0.578	0.397	0.573	0.606	0.704	0.623
F_{TV}	0.793	0.472	0.538	0.621	0.858	0.633	0.621
F_{DIL}	0.888	0.460	0.808	0.699	0.891	0.662	0.607
$F(p^*, 1/2)$	0.918	0.554	0.852	0.913	0.832	0.693	0.699

In this regard, the proposed diffusivity was also tested with abdominal CT scans. Obviously, the image domains were in 3D and according to clinical protocols, the grids were not regular. We observed that the choice of γ as global parameter was not a good one. Specifically, the weak edges were blurred and then some regions of interest were merged. To overcome these drawbacks, several strategies about a local γ are being studied. Our approach is to detect weak edges and sequently apply a local γ that tends to zero for these areas. In the rest of the image, γ could be global and calculated using MAD function as we stated above. These experiments have validated the 3D extension of our filter, even with high values of k. Fig. 10 shows first preliminary results.

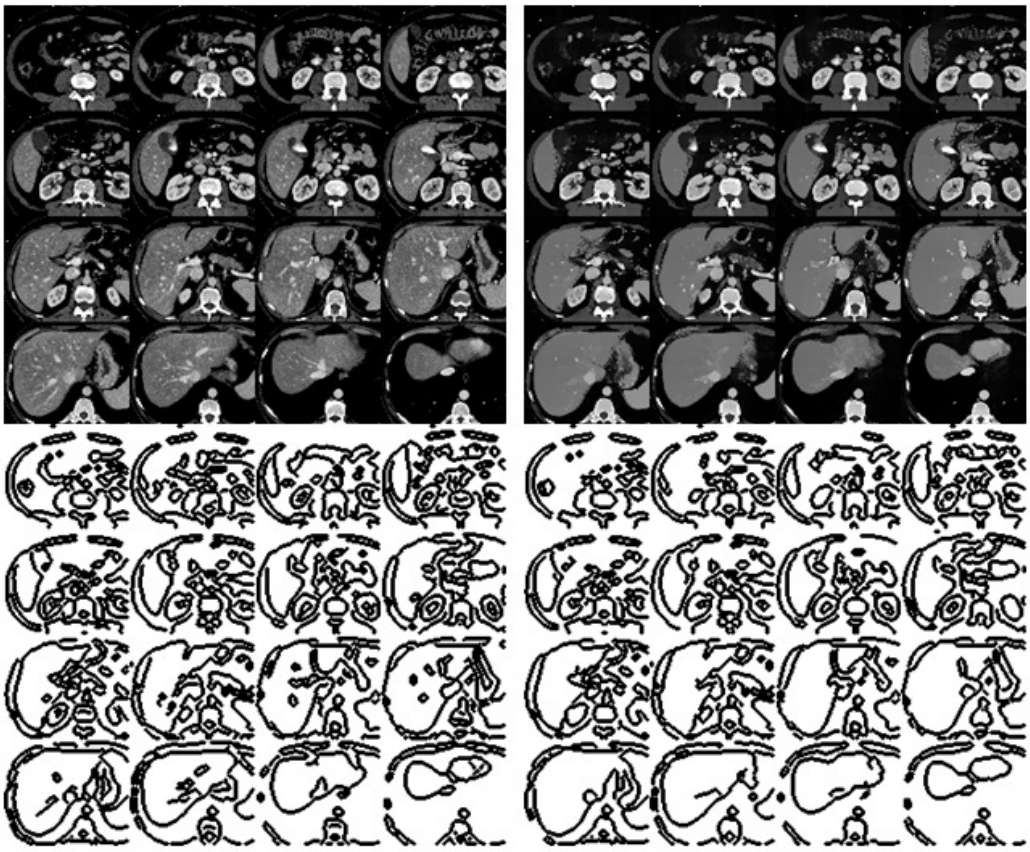


Figure 10: Some slices of the original CT and the corresponding filtered slices in the last iteration. The size of the image is $296 \times 357 \times 67$, pixel spacing is 0.76 *mm* and slice distance is 3 *mm*. The values are $p^* = 10$, $k = 8000$ and $n = 15$.

7 Discussion and Conclusion

In the first part of this article, the intrinsic formulation for nonlinear diffusion equation was developed (5). This allowed us, firstly, to observe clearly the effect of the derivative of the diffusion function in the condition for the edge enhancement, and secondly to establish a theoretical framework for the design of the diffusivity function. Because of this, diffusivity function (16) was proposed, capable of separating the initial image in piecewise constant regions using nonlinear diffusion techniques.

From the numerical point of view, it has been observed that with the proposed diffusivity (16), the Newton's method was unstable, due to the matrix of iteration. Therefore, we have suggested to use the tangential stiffness method given by (29), which leads to the *Picard* or *direct* method (30). In the case of the first iteration we obtain the *semi-implicit method*. Thus, it is shown that with this numerical method we really perform forward diffusion instead of backward diffusion connecting with an edge-preserving process.

The diffusivity (16) has two parameters: γ is related to the slopes at the edges of the regions, while p is related with the speed of the forward diffusion. It also has interesting properties. First, the value of $g\left(\|\nabla u\|^2\right) = 1$ for $\|\nabla u\| < \sqrt{\gamma}$ allows to use the *setting time* t_s as a stopping time. Second, we have decided to set the parameter γ by using the MAD function in the different test images. So, depending on the features of each image and using a set of training images, we can adjust the parameter p as it was stated in the experimental part. This fact allows us to introduce prior knowledge for getting a piecewise constant image using the criterion: $p^* = \arg\max\{F(p, 1/2) : p > 1\}$. Thus, we get good results for images that have some contrast among the regions (see images in Fig. 8: bear, elephant, flower, grave, sheep). Third, we emphasize that we only need to calculate once the *setting time* t_s for one image of the set and it can be used for the rest of the images if they have the same size and when $\boldsymbol{U}^0$ is similar. Fourth, the robustness of our diffusivity (16) leads to the use of a large time step $k \equiv \Delta t$, therefore reducing the number of steps, as it can be observed in the liver image example (Fig 10). This feature is an advantage over the functions of the type appearing in equation (15), for which it is necessary to use increments of time too small. From the last two properties, it follows that the proposed diffusivity provides a nonlinear diffusion process for getting piecewise constant images with a low computational effort. This fact, combined with the no necessity to consider functional spaces to select

the texture and the geometry of the image processing, give us an alternative over variational methods. Finally, as previously mentioned, the diffusivity function was tested for 2D and 3D, with promising results.

It is worth noting that for images where the contrast is lower among different regions, the result is not conclusive (see above images in Fig. 8: scissors, bush). To improve outcomes in these cases we propose, as future work, to use a local value of γ. Fig. 10 shows some initial results of this approach. Anyway, these topics are not discussed in this article because its development and approach make the size of this item excessive. These approaches are also postponed for future work.

Finally, a comparative study with other diffusivities [5, 11, 12, 29] is currently under investigation.

References

[1] Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations (Second Edition). Springer-Verlag (2006)

[2] Aubert, G., Vese, L.: A variational method in image recovery. SIAM Journal on Applied Mathematics **34**(5), 1948–1979 (1997)

[3] Aujol, J.M., Gilboa, G.: Constrained and snr-based solutions for tv-hilbert space image denoising. Journal of Mathematical Imaging and Vision **26**(1), 217–237 (2006)

[4] Aujol, J.M., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition-modeling, algorithms, and parameter selection. International Journal of Computer Vision **67**(1), 111–136 (2006)

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

[6] Brox, T.: From pixels to regions: Partial differential equations in image analysis. Ph.D. Thesis. Mathematical Image Analysis Group; Dep. of Math. and Comp. Science (Saarland University-Saarbrucken-Germany) (2005)

[7] Buades, A., Le, T.M., Morel, J.M., Vese, L.: Fast cartoon + texture image filters. IEEE Transactions on Image Processing **19**(8), 1978–1986 (2010)

[8] Catte, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Applied Mathematics **29**(1), 182–193 (1992)

[9] Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numerische Mathematik **76**(2), 167–188 (1997)

[10] Chan, T.F., Vese, L.: Active contours without edges. IEEE Trasanctions on Image Processing **22**(1), 61–79 (2001)

[11] Chao, S., Tsai, D.: Anisotropic diffusion with generalized diffusion coefficient function for defect detection in low-contrast surface images. Pattern Recognition **43**(5), 1917–1931 (2010)

[12] Chao, S., Tsai, D.: An improved anisotropic diffusion model for detail- and edge-preserving smoothing. Pattern Recognition Letters **31**(13), 2012–2023 (2010)

[13] Criminisini, A.: Image and video editing. URL `http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm`

[14] Dolceta, I.C., Ferreti, R.: Optimal stopping time formulation of adaptative image filtering. Applied Mathematics and Optimization **43**, 245–258 (2001)

[15] Fernandez, J.J., Li, S.: An improved algorithm for anisotropic nonlinear diffusion for denoising cryo-tomograms. Journal of Structural Biology **144**, 152–161 (2003)

[16] Gilboa, G.: Nonlinear scale space with spatially varying stopping time. IEEE Transaction on Pattern Analysis and Machine Intelligence **30**(12), 2175–2187 (2008)

[17] Gilboa, G., Sochen, N., Zeevi, Y.Y.: Forward-and-backward diffusion processes for adaptive image enhancement and denoising. IEEE Trasanctions on Image Processing **11**(7), 689–703 (2002)

[18] Gurtin, M.E.: An Introduction to Continuum Mechanics. Academic Press (1981)

[19] Ilyevsky, A., Turkel, E.: Stopping criteria for anisotropic pdes in image processing. Journal of Scientific Computing **45**, 333–375 (2010)

[20] Kawohl, B.: From mumford-shah to perona-malik in image processing. Mathematical Methods in the Applied Sciences **25**(15), 1803–1814 (2004)

[21] Keeling, S.L., Stollberger, R.: Nonlinear anisotropic diffusion filters for wide range edge sharpening. Inverse Problems **18**, 175–190 (2002)

[22] LeVeque, R.J.: Finite Difference Methods for Ordinary and Partial Differential Equations. SIAM (2007)

[23] Martin, D., Fwolkes, C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transaction on Pattern Analysis and Machine Intelligence **26**(1), 1–20 (2004)

[24] Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations. American Mathematical Society (University Lecture Series) **22** (2001)

[25] Mrazek, P., Navara, M.: Selection of optimal stopping time for nonlinear diffusion filtering. International Journal of Computer Vision **52**(2/3), 189–203 (2003)

[26] Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communitcations on Pure and Applied Mathematics **42**, 577–685 (1989)

[27] Ortega, J.M.: Numerical Analysis (A Second Course). SIAM, Philadelfia (1990)

[28] Owen, D.R.J., Hinton, E.: Finite Elements in Plasticity. Pineridge Press Limited, Swansea (1980)

[29] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transaction on pattern analysis and machine intelligence **12**(7), 629–639 (1990)

[30] Platero, C., Sanguino, J., Velasco, O.: Nonlinear diffusion filters without parameters for image segmentation. Lecture Notes in Computer Science (13th International Conference, CAIP 2009) **5702**, 517–524 (2009)

[31] Reddy, J.N.: Nonlinear Finite Element Analysis. Oxford University Press, New York (2004)

[32] Rifkah, E., Amer, A.: Fast automated stopping-time and edge-strength estimation for anisotropic diffusion. In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 933–936. IEEE (2008)

[33] Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D. **60**, 259–268 (1992)

[34] Solo, V.: Automatic stopping criterion for anisotropic diffusion. In: Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, vol. 6, pp. 3929–3932. IEEE (2001)

[35] Tsurkov, V.I.: An analytical model of edge protection under noise suppression by anisotropic diffusion. Journal of Computer and Systems Sciences International **39**(3), 437–440 (2000)

[36] Van Rijsbergen, C.: Information Retrieval. Second ed. Dept. of Computer Science, Univ. of Glasgow (1979)

[37] Vanhamel, I., Mihai, C., Sahli, H., Katatzis, A., Pratikakis, I.: Scale selection for compact scale-space representation of vector-valued images. International Journal of Computer Vision **84**(2), 194–204 (2009)

[38] Varga, R.S.: Matrix Iterative Analysis. Springer (1991)

[39] Vese, L.: A study in the bv space of a denoising-deblurring variational problem. Applied Mathematics and Optimization (44), 131–161 (2001)

[40] Vese, L., Osher, S.: Modeling textures with total variation minimization and oscillating patterns in image processing. Journal of Scientific Computing **44**(2), 131–161 (2003)

[41] Wang, Y., Ruiqing, N., Liangpei, Z., Wu, K., Sahli, H.: A scale-based forward-and-backward diffusion process for adaptative image enhancement and denoising. EURASIP Journal on Advances in Signal Processing **2011:22** (2011)

[42] Wang, Y., Zhang, L., Li, P.: Local variance-controlled forward-and-backward diffusion for image enhancement and noise reduction. Image Processing, IEEE Transactions on **16**(7), 1854–1864 (2007)

[43] Weickert, J.: Anisotropic Diffusion in Image Processing. B. G. Teubner, Stuttgart (Deutschland) (1998)

[44] Weickert, J.: Coherence-enhancing diffusion of colour images. Image and Vision Computing **17**(3), 201–212 (1999)

[45] Weickert, J., Benhamouda, B.: A semidiscrete nonlinear scale-space theory and its relation to the perona-malik paradox. Advances in Computer Vision pp. 1–10 (1997)

[46] Weickert, J., ter Haar Romeny, B.M., Viergever, M.A.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trasanctions on Image Processing **7**(3), 398–410 (1998)

[47] Welk, M., Gilboa, G., Weickert, J.: Theoretical foundations for discrete forward-and-backward diffusion filtering. Scale Space and Variational Methods in Computer Vision pp. 527–538 (2009)