Pattern Recognition 47 (2014) 270–281

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

On the impact of anisotropic diffusion on edge detection

C. Lopez-Molina [a,b,*], M. Galar [a], H. Bustince [a], B. De Baets [b]

[a] Dpto. Automatica y Computacion, Universidad Publica de Navarra, 31006 Pamplona, Spain
[b] KERMIT, Dept. of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure links 653, 9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form
19 June 2013
Accepted 4 July 2013
Available online 26 July 2013

Keywords:
Image regularization
Content-aware smoothing
Anisotropic Diffusion
Edge detection
Quantitative evaluation

ABSTRACT

Content-aware, edge-preserving smoothing techniques have gained visibility in recent years. However, they have had a rather limited impact on the edge detection literature compared to content-unaware (linear) techniques, often based on Gaussian filters. In this work, we focus on Anisotropic Diffusion, covering its initial definition by Perona and Malik and subsequent extensions. A visual case study is used to illustrate their features. We perform a quantitative evaluation of the performance of the Canny method for edge detection when substituting linear Gaussian smoothing filters by Anisotropic Diffusion.

1. Introduction

Edge detection is often based on the analysis of intensity differences within pixel neighbourhoods. Intuitively, this leads to the computation of the partial derivatives or the Laplacian of a signal [1,2]. However, due to the discrete nature of the data, the classical concept of differentiation cannot be applied. The ill-posedness of the gradient computation is partially solved by regularizing (smoothing) the image. The simplest and most common way to do so is content-unaware smoothing (CUS), which is usually performed using two-dimensional Gaussian filters [3,4]. CUS techniques remove noise and image imperfections at the cost of blurring the edges [5]. For this reason, there is a need for smoothing techniques that regularize the pixel intensities within the image objects, while preserving (or even sharpening) their boundaries. As enunciated by Monteil and Beghdadi [6], the goal is to combine *a low pass filtering in homogeneous regions and a sharpening effect in transition regions*.

Content-aware smoothing (CAS) techniques adjust their behaviour based upon local features. One of the most relevant CAS techniques is *Anisotropic Diffusion* (AD), initially proposed by Perona and Malik [5], who formulate the smoothing problem in terms of heat diffusion. They were aiming at a process in which the image properties (in this case, intensity) would spread inside the objects but not across their boundaries. In this way, AD allows heat (intensity) diffusion inside the objects, inhibiting the heat transfer across the edges. Following [5], different approaches to AD have been explored, incorporating notions from statistics [7], fuzzy logic [8], and photometry [9], among others. Moreover, AD has proven valid not only for image regularization, but also as inpainting method in the reconstruction of missing (deleted) information [10–12].

Despite the variety of CAS techniques (either based on AD or not), most of the works on edge detection still use linear filtering [13,14] or do not even regularize the image [15,16]. One of the reasons is the lack of quantitative comparisons confronting CUS and CAS techniques. Indeed, most of the works proposing CAS for edge detection contain sample images where the improvements can be observed, but do not quantify the results. The few works using objective measures [17,18] are related to noise removal. Note that the lack of quantitative comparisons is not exclusive to CAS proposals, but rather endemic to the edge detection field [19], to such an extent that Papari and Petkov state that quantitative comparison is *absent from most of the works in edge detection* [20].

In this work we take AD as a paramount example of CAS, and we list three objectives:

(a) to briefly analyze the technical specifications of the relevant AD methods for edge detection,
(b) to illustrate the effect of such AD methods on natural images, and
(c) to quantify the improvement that can be achieved by a typical edge detector if replacing Gaussian linear filtering by any of the aforementioned AD methods.

We approach these tasks from a practical point of view. First, we review the most relevant AD methods in the literature, pointing out their connection with the original proposal by Perona and

* Corresponding author at: Dpto. Automatica y Computacion, Universidad Publica de Navarra, 31006 Pamplona, Spain. Tel.: +34 948 166 331; fax: +34 948 168 924.
E-mail addresses: carlos.lopez.molina@gmail.com, carlos.lopez@unavarra.es (C. Lopez-Molina), mikel.galar@unavarra.es (M. Galar), bustince@unavarra.es (H. Bustince), bernard.debaets@ugent.be (B. De Baets).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.07.009

Malik, as well as listing their parameters and settings. Second, we apply different AD methods to a natural image, in order to observe and compare the changes the image undergoes. Third, we perform a quantitative comparison of the results produced by the Canny method for edge detection [3] when combining it with Gaussian Linear Filtering (GLF) and different AD methods. In this work we consider the scalar approach to AD [5,7], as well as approaches based on vectorial gradients [21–23] and speckle models [18,24].

The remainder of this work is organized as follows. Section 2 contains a review of the literature. In Section 3 we analyze how different AD methods use local information to drive the diffusion process. Section 4 includes a visual example of the transformation of an image when different AD methods are applied. To conclude, the experimental results and conclusions are presented in Sections 5 and 6, respectively.

2. Anisotropic Diffusion

Anisotropic Diffusion is an example of *scale-space* image processing [5,22,25,26]. It stems from the application of the heat diffusion equation to digital images. Considering an image I, the heat flux ϕ within the image is given by Fick's equation

$$\phi = -D \cdot \nabla I, \tag{1}$$

where the relationship between the gradient ∇I and the actual flux j is expressed by a positive definite, symmetric matrix D referred to as *diffusion tensor*. Since the diffusion process does not alter the overall energy in the image, its local variation is driven by $\delta_t I = -\mathrm{div}\phi$ [22], where div stands for the divergence operator. Hence, we have that

$$\delta_t I = \mathrm{div}(D \cdot \nabla I) \tag{2}$$

expresses the energy (heat) variation at every position in the image. The abstraction in Eq. (2) has been embodied in different ways, leading to the creation of a wide family of AD methods. In this work we focus on five such methods, which we consider representative:

(i) *Perona–Malik Anisotropic Diffusion*: The first reference to the application of AD to digital images is due to Perona and Malik [5]. The authors consider the diffusion process on an image I to be modelled by

$$\delta_t I = \mathrm{div}(g(|\nabla I|^2) \cdot \nabla I), \tag{3}$$

where the function g is an edge-stopping, decreasing function weighing the conductivity of the image depending upon the Euclidean magnitude of the gradient. However, they propose a simpler, discrete scheme based on the transfer of energy between each pixel and its four direct neighbours. Despite its simplicity, this scheme has some interesting properties such as the preservation of the energy or the fact that it does not create new (local) maxima or minima. We will refer to this method as Perona–Malik AD (PMAD).

Almost at the same time, Saint-Marc et al. [27] introduced the notion of adaptive smoothing, which turns out to be an alternative implementation of PMAD (see [28] for a historical perspective). Catté et al. [29] proposed to compute the gradients on a regularized version of the image I, such as $I_\sigma = G_\sigma * I$, where G_σ represents a Gaussian filter with standard deviation σ and $*$ is the convolution operator. In this way, the authors claim to solve the *inconsistencies* of PMAD, more specifically the fact that small changes in the initial image *can produce divergent solutions* [29]. Subsequently, many authors have revisited PMAD, mainly focusing on the instability of the process and the staircasing effect (creation of non-existing step edges) [30–32].

(ii) *Diffusion Tensor-based Anisotropic Diffusion*: Cottet and Germain [23] argue against the use of the name AD in [5], since PMAD makes use of scalar conductivity values instead of diffusion tensors. From Eq. (2) we obtain isotropic and non-linear isotropic diffusion using

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad D = \begin{pmatrix} g(|\nabla I|) & 0 \\ 0 & g(|\nabla I|) \end{pmatrix}, \tag{4}$$

respectively [33]. In order to obtain an anisotropic behaviour, we must use the direction of ∇I, not only its magnitude. Diffusion Tensor-based AD (DTAD) is based on Eq. (2), where the eigenvectors of D are $v_1 \| \nabla I_\sigma$ and $v_2 \perp \nabla I_\sigma$. In order not to smoothen across the edges, the eigenvalues are set to $\lambda_1 = g(|\nabla I_\sigma|^2)$ and $\lambda_2 = 1$ [22].

(iii) *Structure Tensor-based Anisotropic Diffusion*: A *structure tensor* (J_ρ) is a symmetric matrix associated with a gradient ∇I_σ constructed as

$$J_\rho(\nabla I_\sigma) = G_\rho * (\nabla I_\sigma \otimes \nabla I_\sigma). \tag{5}$$

The structure tensor is useful to characterize the underlying structure and features of the image, not only the local intensity variations. It differs from the diffusion tensor in the fact that it captures the orientation of the intensity change, not its direction. The integration parameter ρ has to be set according to the size of the underlying image structure [22]. Note that σ and ρ play similar roles, although they have different purposes. The parameter σ aims to remove (or reduce the impact of) noise and imperfections in the image, and hence is related to the contamination of the image, not to its content. Alternatively, ρ takes values according to the size of the structures and object silhouettes one wants to preserve. When using structure tensors, we replace the matrix D in Eq. (2) by a matrix with the same eigenvectors as J_ρ and eigenvalues

$$\lambda_1(\mu_1) = g(\mu_1) \quad \text{and} \quad \lambda_2 = 1 \tag{6}$$

where μ_1 is the greatest eigenvalue of J_ρ. The AD method based on structure tensors is referred to as STAD.

(iv) *Coherence Enhancing Anisotropic Diffusion*: The purpose of Coherence Enhancing AD (CEAD) [34] is intrinsically different from that of the other AD methods. In this approach the smoothing is not prevented across the edges, but empowered along them instead. That is, there is heat transfer in edge regions, but it takes place along the edges. In this approach the matrix D in (2) is replaced by a matrix with the same eigenvectors μ_1 and μ_2 as the structure tensor J_ρ and eigenvalues

$$\lambda_1 = \alpha \quad \text{and} \quad \lambda_2 = \begin{cases} \alpha, & \text{if } \mu_1 = \mu_2 \\ \alpha + (1-\alpha) \cdot g(\mu_1 - \mu_2), & \text{otherwise} \end{cases}, \tag{7}$$

where $\alpha \in]0,1[$ is a regularization parameter that ensures a certain amount of diffusion to occur in situations with isotropic intensity change, and takes a positive value close to 0 [21].

(v) *Speckle Reducing Anisotropic Diffusion*: Some authors have presented AD methods whose local behaviour is not determined by gradients. A relevant example is the methods based on noise estimation, such as Speckle Reducing Anisotropic Diffusion (SRAD), introduced by Yu and Acton [18]. Speckle noise is a very common type of contamination in digital images. It is manifested as random impulse noise introducing normally distributed variations in the pixels of an image. SRAD intends to regularize the image in the presence of speckle, while not carrying out regularization in homogeneous regions or edge regions. In SRAD

Table 1
List of AD methods considered in this work.

Method (Abr.)	Behaviour	Diffusivity	Year
Perona–Malik AD (PMAD)	Edge preserving	Scalar	1990
Diffusion Tensor-based AD (DTAD)	Edge preserving	Vectorial, diffusion tensor	1993
Structure Tensor-based AD (STAD)	Edge preserving	Vectorial, structure tensor	1998
Coherence Enhancing AD (CEAD)	Structure reinforcing	Vectorial, structure tensor	1999
Detail-preserving AD (DPAD)	Noise removing	Scalar, local statistics	2006

the *conductivity* (diffusion coefficient) is generated based on local statistics of each pixel. In the original formulation, we have

$$\delta_t I = \mathrm{div}(c(q(I)) \cdot \nabla I), \tag{8}$$

where q is a local measure of the speckle, and c is a function modulating the amount of diffusion that takes place.
Although SRAD has different goals than other AD methods, they share some features. The role of the function c in SRAD is very similar to that of the conductivity function g in the previous AD methods. In this work we will use a more evolved version of SRAD, namely Detail-Preserving AD (DPAD), which is also based on Eq. (8), but differs in the extraction of local features.

Other works have been published analyzing the relationships between AD and other fields (such as robust statistics [7], multigrid techniques [35,36] or, in a more general way, non-linear filters [37]), introducing new diffusion schemes [8,38] or studying the properties (mainly stability) of the diffusion process [21,27,30,39]. Also methodologies for content-aware smoothing other than AD have been introduced in the literature, e.g. bilateral filtering [40,41].

In Table 1 the methods considered in this comparison are summarized, together with their characteristics.

3. Extraction of local information for diffusion processes

Anisotropic Diffusion, as any other CAS technique, modifies its behaviour at every position in the image based on local measurements. Hence, the accuracy of these measurements is of paramount importance to obtain trustworthy results. In this section we analyze the use of local information in the AD methods reviewed in Section 2.

3.1. Gradient computation

Most of the AD methods in the literature are based on gradient measurements, since gradients with large magnitudes are typically seen as a local manifestation of edges [2].[1] However, gradient computation on digital images is an ill-posed problem, due to the discrete nature of the data [43]. The accuracy and reliability of gradient computation techniques have been widely studied in the literature [3,4,20,44]. Although the number of edges and their exact position cannot be known, the gradients provide a *current best estimate of the location of the boundaries* [5].

PMAD does not compute the gradient at each pixel, but the difference between its intensity and that of each of the four direct neighbours. This scheme radically simplifies the problem, but is prone to generate problems in noisy images [29]. In this work, we keep it in this way to maintain the original proposal.

When using DTAD, STAD or CEAD, we need to compute ∇I. In this work we use the Canny operators for feature extraction, a common choice in the literature [3,4]. By using these operators, the gradient at a given pixel is computed as

$$\nabla I = (I * (-G_{\sigma,x}), I * (-G_{\sigma,y})) \tag{9}$$

where $G_{\sigma,x}$ and $G_{\sigma,y}$ represent the first derivatives of a two-dimensional Gaussian filter with standard deviation σ (i.e. G_σ). Note that the negation of the Gaussian derivative filters is used for the gradient to be in accordance to the classical coordinate system (i.e. increasing to the right and upward). In the remainder of this work the standard deviation of the filter used for computing gradients is referred to as σ_f.

3.2. Image regularization

A simple way to overcome the ill-posed nature of gradient computation is to use a regularized version of the image. This is usually done with Gaussian filters, since they have signal conservation properties of great interest [25,26,45]. In this way, the gradient computation (Eq. (9)) is performed on $I_\sigma = I * G_\sigma$ instead of on I. The problem of adequately selecting the value of σ is recurrent in the literature, and is deeply studied in the Gaussian Scale-Space [4,25,26].[2] In the remainder of this work the standard deviation of the filter used to regularize the image prior to feature extraction is referred to as σ_r.

In order to illustrate the importance of σ_r, we display in Fig. 1 the Euclidean magnitude of the gradients in image 101085 of the Berkeley Segmentation Dataset (BSDS) [46] after smoothing the image with different Gaussian filters ($|\nabla I_{\sigma_r}|$). Low values of σ_r result in large values of $|\nabla I_{\sigma_r}|$ in regions with high frequency artifacts (e.g. in the grass regions in Fig. 1(b) and (c)). This causes highly inhibited diffusion of unimportant features (when using DTAD and STAD) or large structural reinforcement of them (when using CEAD). Alternatively, high values of σ_r tend to blur the edges. This fact is especially clear in the regions with high-contrast edges, such as the top left part of the leftmost totem in Fig. 1(f). In this case, the edge preservation or reinforcement suffers from low spatial accuracy. This trade-off is the same as the one enunciated by Canny between spatial accuracy and good discrimination of true edges [3,26,47].

3.3. Local feature regularization

In addition to image regularization, when applying STAD and CEAD we also have to consider the spatial regularization of the structure tensors, which uses the integration parameter ρ. Large values of ρ nto only help overcoming small imperfections in large structures, but also hinder the adequate interpretation of small objects.

In Fig. 2 we display both the value of the greatest eigenvalue of $J_\rho(\nabla I_{2.0})$ and its difference with the second eigenvalue using different values of ρ. The upper row (μ_1) represents the amount of intensity change in the *main* direction, while the lower row ($\mu_1 - \mu_2$) illustrates the anisotropy of the intensity change. When

[1] Some applications on special imagery rely on descriptors other than the gradient, e.g. texture descriptors [42].

[2] The GSS provides a framework for edge characterization with increasing Gaussian smoothing. However, it does not intend to produce a multi-scale gradient characterization or computation.

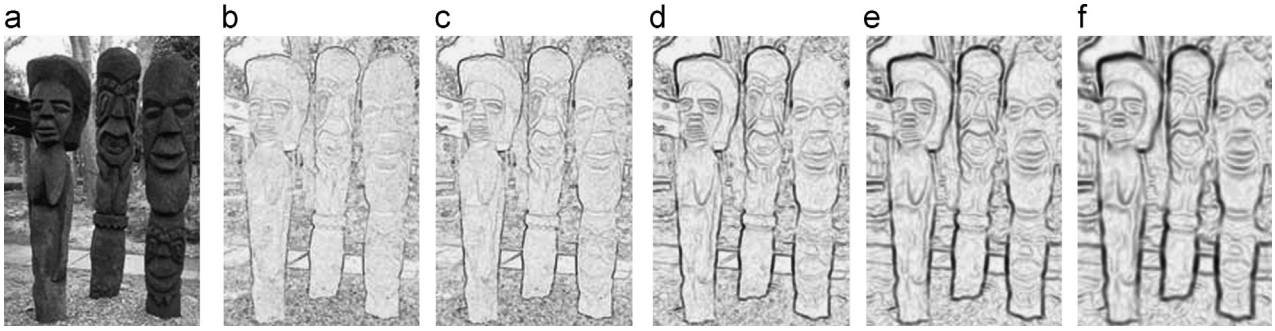

Fig. 1. Euclidean magnitude of the gradients computed with the Canny operators for feature extraction ($-G_{1.0,x}$, $-G_{1.0,y}$) on Gaussian-regularized versions of an image. (a) Original image, (b) $|\nabla I_{0.25}|$, (c) $|\nabla I_{0.5}|$, (d) $|\nabla I_{1.0}|$, (e) $|\nabla I_{1.5}|$ and (f) $|\nabla I_{2.0}|$.

the intensity change occurs with similar strength in orthonormal directions we have $\mu_1 \approx \mu_2$, indicating that it is not an edge, but a noisy pixel instead. Two different situations involving intensity changes are observed in Fig. 2: edges and textures. In edge regions, such as at the top left side of the leftmost totem, large values for both μ_1 and $\mu_1 - \mu_2$ are obtained, since the intensity change in the orthonormal direction (μ_2) is minimal. In regions with textures (such as the grass region) only large values of the integration parameter ρ reveal the absence of an edge.

Small values of ρ do not allow to distinguish between true edge and texture regions, since they both appear to contain true edges ($\mu_1 \gg \mu_2$). Therefore, larger values of ρ are needed to be able to distinguish between the edges that are to be enhanced/preserved (silhouettes) from those that are to be smoothed (textures). Logically, this leads to a low spatial accuracy and poor recognition of small objects. As in Section 3.2, this poses a scale-determination problem.

Since there is no consensus on how to set the values of σ_r and ρ, in further sections we will use several combinations of them.

3.4. Speckle estimation

Among the methods presented in Section 2, SRAD and DPAD are the only ones not relying on gradient computations, using instead estimations of the presence of speckle. SRAD uses at each position p

$$q(p) = \frac{\sqrt{\frac{1}{|n(p)|-1}\sum_{p' \in n(p)} (I(p') - \overline{n_p})^2}}{\overline{n_p}}, \tag{10}$$

where $n(p)$ is a neighbourhood centered on p, with average intensity $\overline{n_p}$. Yu and Acton [18] use

$$c(q) = \frac{1}{1 + \frac{q^2 - \theta^2}{\theta^2 \cdot (1+\theta^2)}} \quad \text{or} \quad c(q) = e^{(q^2-\theta^2)/(\theta^2 \cdot (1+\theta^2))}, \tag{11}$$

where θ is a parameter representing the strength of the speckle in the image. In order to calculate the value of θ, they propose to use either local statistics on a homogeneous subregion of the image (which is hardly applicable in unsupervised image processing) or a decreasing function blindly forecasting the reduction of the speckle [18].

The choice of the coefficients q and θ in SRAD, as well as the function c, was criticized in [24], in which the authors introduce a different function c

$$c(q) = \frac{1 + q^{-2}}{1 + \theta^{-2}}. \tag{12}$$

They also propose and test a new estimation technique for the parameter θ, namely

$$\theta = \text{Median}_{p \in I}(q^2). \tag{13}$$

The AD method driven by Eqs. (8), (12) and (13) is referred to as Detail-Preserving AD (DPAD) [24]. SRAD and DPAD have been widely applied to the processing of medical imagery, which is especially prone to speckle (see, for example, [48]). Some contributions have subsequently evolved from SRAD, such as the Oriented SRAD (OSRAD), using a vectorial representation of the conductivity [49].

3.5. Conductivity function

The conductivity function g modulates the amount of diffusion depending upon a local characteristic. The first proposals in [5] are

$$g_1(m) = e^{-(m/K)^2} \quad \text{and} \quad g_2(m) = \frac{1}{1 + \left(\frac{m}{K}\right)^2}, \tag{14}$$

where m is the gradient magnitude[3] and K is a threshold to distinguish those gradients due to true edges from those due to spurious objects (see Fig. 3). Black et al. perform a detailed study of conductivity functions (also proposing to use Tukey's biweight function or Huber's minimax estimator) [7], while Whitaker and Pizer analyze their interpretation and adjustments [31].

Although SRAD (as well as DPAD) does not make use of conductivity functions, the function c plays a similar role. However, it is derived from the underlying model of well-known speckle filters [24], not from a proper feature-to-conductivity mapping.

3.6. Numerical scheme and time discretization

A diffusion problem is defined by the initial state (in this case the original image) and a diffusion scheme (reviewed in Section 2), which defines the evolution of the image as time progresses. With this information the diffusion process is, theoretically, completely defined. However, diffusion processes happen in continuous environments with infinite precision, which is not the case for digital images. The problems due to the discrete nature of the image are usually tackled by regularizing the image (I_σ), but there is also a need to evolve the image using discrete time steps, that is, to sequentially generate evolving versions of the image ($I^{[t+\Delta t]}$) from a given time-preceding one ($I^{[t]}$).

The explicit numerical schemes are the simplest option, as well as the most widely used in the literature [21]. These schemes are called *explicit* because each $I^{[t+\Delta t]}$ is computed directly from $I^{[t]}$.

[3] In the original formulation of PMAD m represents the difference between a pixel and each of its neighbours, which is not properly a gradient [5]. Subsequent proposals for AD use the gradient magnitude for m.

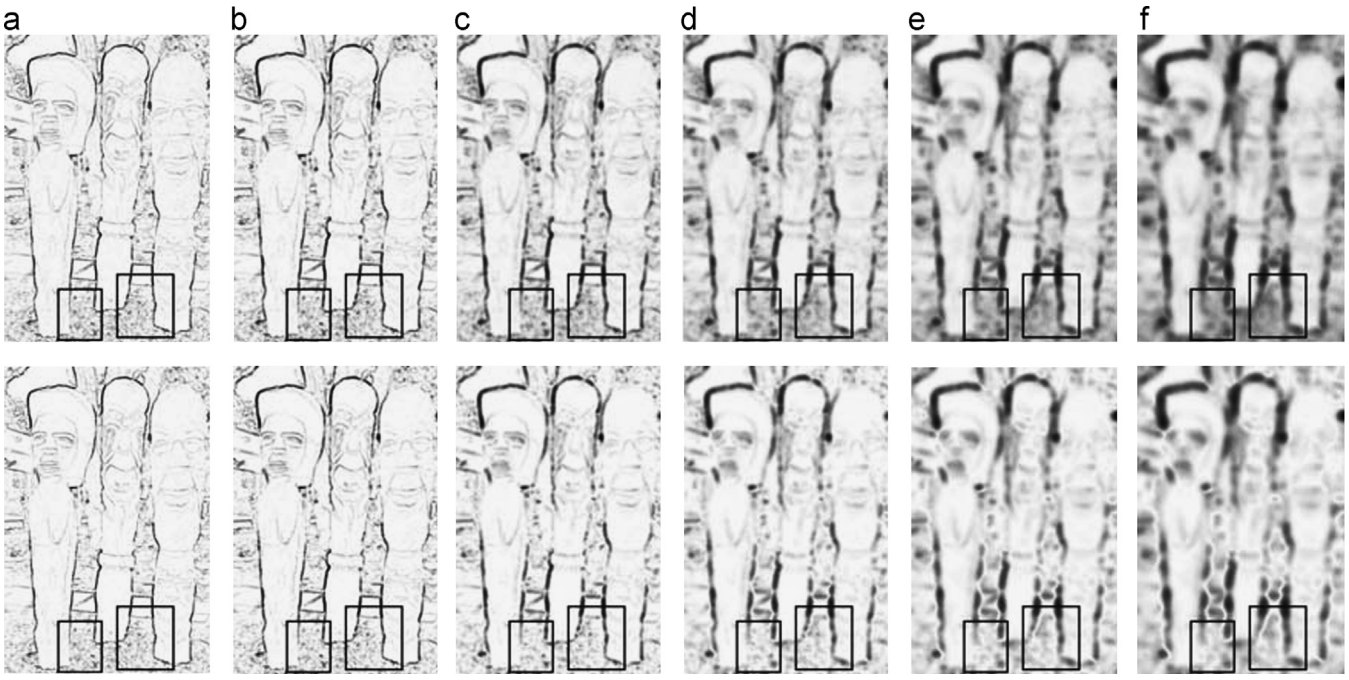

Fig. 2. Value of the greatest eigenvalue (first row) and difference of the eigenvalues (second row) of the diffusion tensor D generated after a structure tensor $J_\rho(\nabla I_{2.0})$ with different values of ρ, and the image in Fig. 1(a). (a) $\rho=0.5$, (b) $\rho=1.0$, (c) $\rho=2.0$, (d) $\rho=3.0$, (e) $\rho=4.0$ and (f) $\rho=5.0$.

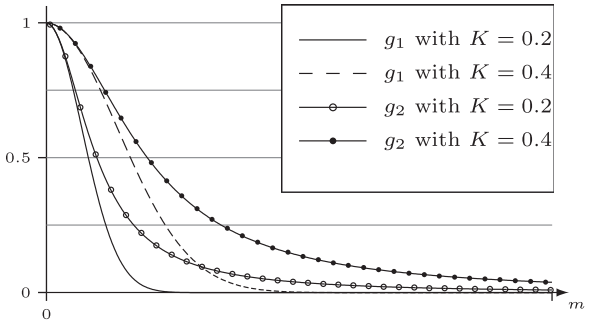

Fig. 3. Conductivity functions g proposed by Perona and Malik.

Their main drawback is that they demand very small time steps in order to ensure stability. Variants of the explicit schemes have been developed, including semi-implicit discretization (which implies the resolution of a system of equations at each time t) or the Additive Operator Splitting (AOS) [50,51]. These schemes outperform the explicit ones in terms of efficiency, since they permit larger time steps and, consequently, require less images to be computed to reach a given diffusion time t. A comprehensive study of numerical schemes for AD is presented in [22].

In this work we have used an explicit implementation of the different AD methods, which forces us to perform small time steps (Δt). The reason to do so is that we evaluate the usefulness of AD methods for edge detection, regardless of their efficiency. Moreover, by restricting to explicit schemes we can use the original PMAD.[4]

4. A case study for AD methods

AD has resulted in very different methods, each having different purposes. In this section we analyze the effect of different AD methods on a natural image (Fig. 1(a)) that contains most of the relevant artifacts in the edge detection problem: sharp edges (top of the leftmost totem), vanishing edges (left side of the face of that same totem), and textures with different coarseness and strength (those in the grass or the body of the totems). In this case study we include the AD methods in Table 1. We only use the conductivity function g_2, since the space is limited and other functions did not result in significant differences. For PMAD or DTAD, K is calculated using the Rosin method for thresholding on the histogram of Euclidean magnitudes of ∇I_σ [52], which is computed as in Eq. (9), with $\sigma_f=1.0$. In the case of CEAD or STAD, K is set using the Rosin method on the histogram of absolute differences $|\mu_1-\mu_2|$, with μ_1 and μ_2 being the eigenvalues of J_ρ [22]. We observe very different results for the different AD methods:

(i) PMAD is known to suffer from different problems, mainly staircasing (leading to the creation of new edges) and severe overblurring of smooth edges (potentially eliminating them). In Fig. 4 we show the results obtained by PMAD with different diffusion times. First, we observe that sharp edges are preserved, while the textures are quickly removed. Some of the coarse textures (such as those in the floor) are initially taken as objects, and hence preserved. When $t=50$, the staircasing effect appears in some regions of the background, specially in the grass behind the totems (see Fig. 4(c)). This effect is emphasized as diffusion time progresses, giving rise to well-defined edges in positions with low contrast in the original image (Fig. 4(e)). The overblurring effect takes more time to become visible, but is observable. For example, the face of the leftmost totem is fused with the tree, while the bodies of the right and center ones are *bridged* by the objects in the background (Fig. 4(f)).
When $t=500$ all the textures and irrelevant objects are removed, whereas a very few meaningful silhouettes are destroyed. Nonetheless, PMAD fails to regularize some regions of the background.

(ii) DTAD is a variant of PMAD, using a vectorial representation of the diffusivity at each position of the image. In Fig. 5 we show

[4] Note that in PMAD the authors use an explicit scheme where the role of Δt is taken by a parameter $\lambda\in]0,0.25]$ [5]. In this work we set $\lambda=\Delta t$.

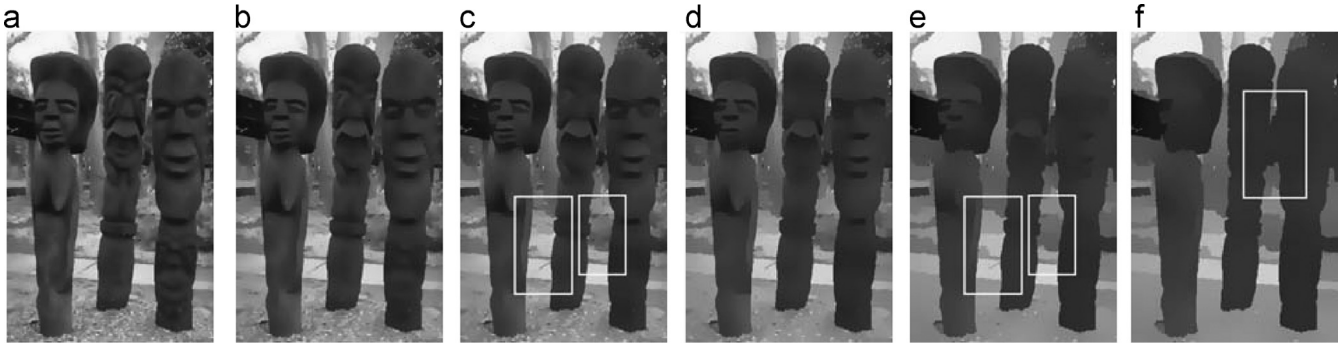

Fig. 4. Image 101085 of the BSDS applying PMAD with different diffusion times t. The parameters $\sigma_r = 0.5$ and $\Delta t = 0.1$ are fixed. (a) $t=10$, (b) $t=25$, (c) $t=50$, (d) $t=100$, (e) $t=300$, and (f) $t=500$.

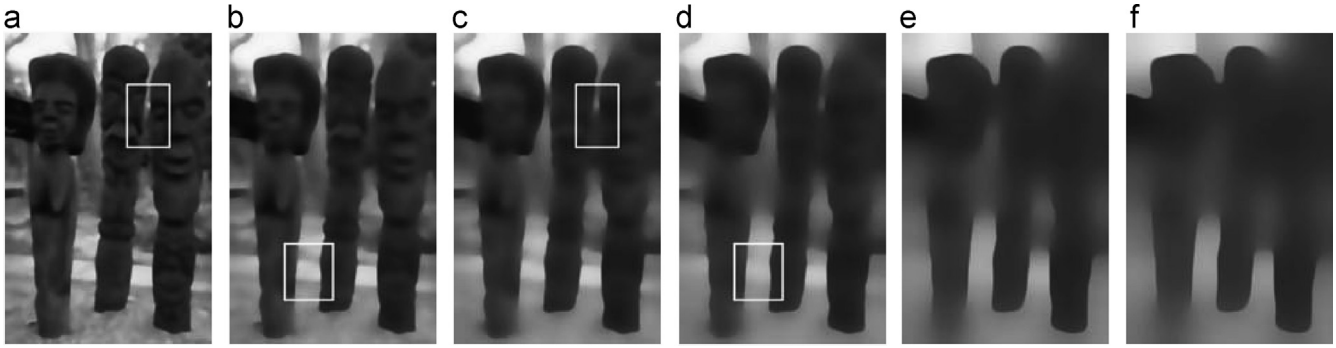

Fig. 5. Image 101085 of the BSDS applying DTAD with different diffusion times t. The parameters $\sigma_r = 0.5$, $\sigma_f = 1.0$ and $\Delta t = 0.1$ are fixed. (a) $t=10$, (b) $t=25$, (c) $t=50$, (d) $t=100$, (e) $t=300$ and (f) $t=500$.

the same sequence of images as displayed in Fig. 4, applying DTAD instead of PMAD. We observe that DTAD produces faster diffusion leading to a quicker destruction of most of the structures in the image. When $t=300$ the objects in the original image are hardly recognizable. We also observe that the textures quickly vanish, being removed at $t=25$. The sharp edges are initially preserved, but most of them eventually fade away.

In Fig. 5 we find two different scenarios under which the edges of the image disappear, which we refer to as *heat leak* and *intra-region compensation*. The schematic representation of both scenarios is given in Fig. 6. In the case of *heat leak* a high-contrast edge is progressively destroyed because of a low contrast region. An example of this scenario can be seen in Fig. 5(c) and (f). The second edge-destructing scenario is the *intra-region compensation*, in which two regions with high-contrast boundaries reach similar heat levels after internal (and independent) diffusion processes. After such diffusion processes, the contrast at the boundary of the regions disappears. This scenario is highlighted in Fig. 5 (a) and (d). The difference between both scenarios is that in the former there is a connection between both regions, probably due to shades or bad illumination. However, in the latter there is a reduction of the contrast at the boundary, despite the absence of diffusion across the edge. Both scenarios severely compromise the quality of the results obtained by DTAD, although they are perfectly explainable from the diffusion scheme. PMAD is based on similar concepts, but it is affected in a different way because of the staircasing effect, which produces new, non-meaningful edges that act as structural barriers preventing edge-destructing diffusion from happening.

(iii) STAD has the same objective as DTAD, but uses the structure tensor for representing the intensity variations at each pixel. The results by STAD with $\rho \in \{1,2\}$ (Fig. 7) are very similar to those by DTAD. The main differences between them arise in texturized regions. In the grass region, for example, the structure tensor (obtained from spatially regularized gradients) allows for a faster diffusion and, consequently, quicker regularization. In this case, when $t=10$, STAD has already regularized the grass region, whereas DTAD needed about 50 time steps. The regularizing effect of the structure tensor has positive and negative effects, but none of them avoids the scenarios in Fig. 6.

(iv) CEAD has a different objective than DTAD and STAD, since it aims to enhance the contrast at the boundary regions. In Fig. 8 we show the results of CEAD, with $\rho \in \{1,2\}$. The results are radically different from the ones by PMAD, DTAD or STAD, since the modifications of the original image are lighter, and mostly restrict to those regions where the most evident boundaries occur. We observe an emphasizing effect of the lines and edges, either meaningful or not. For example, the totem silhouettes are generally reinforced, producing a glowy effect at the edge regions. This visual effect is due to the enhancement of the local contrast at these positions of the image. CEAD needs more time to remove the textures, since it initially reinforces their underlying structure. Besides, some structures are propagated and, as a consequence, merged, thereby creating new objects (see Fig. 8(f)). In the same manner as in DTAD/STAD, this is inherent to the diffusion scheme, since propagation of meaningful edges is coupled to propagation of irrelevant structural features. Note that larger values of ρ lead to faster diffusion and help dealing with the textures, at the cost of losing spatial accuracy.

(v) DPAD is different from the other AD methods, in the sense that the diffusion is not driven by edge features. However, it also aims to ease edge detection, since the noise removal facilitates the discrimination of the real edges from

the spurious responses. Fig. 9 shows the images obtained by applying DPAD under the same conditions used for the other AD methods. First, it can be seen that the alterations in the image are visually very light. Indeed, they restrict to the regularization of some textures in the image whose underlying structure is close to that of the speckle. DPAD is not able to remove the coarse textures in the grass or the top of the trees. We observe that the benefits are limited, but the destruction of the contents of the image is null. Hence, it is an extremely conservative approach that provides little improvement of the image, but is also likely to preserve all of the meaningful information.

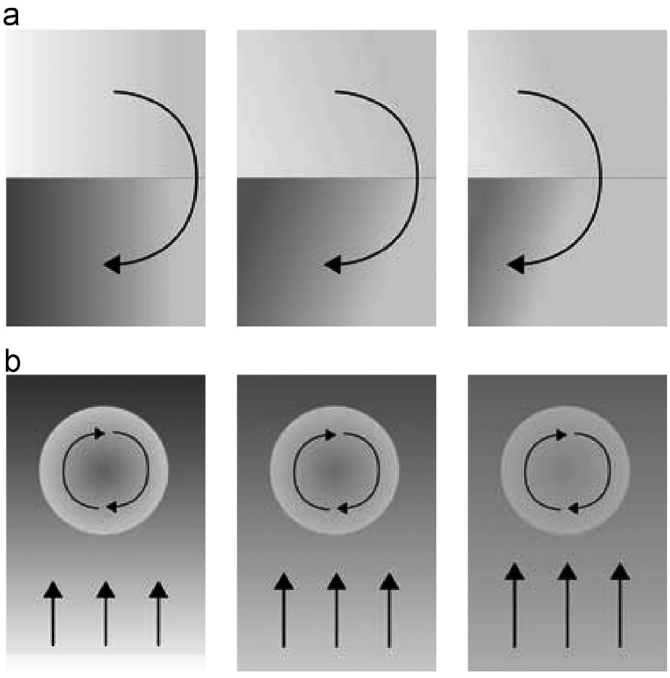

Fig. 6. Edge-destructing scenarios for edge-preserving AD methods. The arrows represent the diffusion fluxes. (a) Heat leak scenario and (b) Intra-region compensation scenario.

From this example two different conclusions can be drawn. First, we observe that the differences between the AD methods are very large. Different diffusion schemes lead to very different final images, which are in general consistent with their goals, as reviewed in Section 2. Second, we have that, sometimes, the conditions of the image lead to unexpected results. Of course, parameter settings other than those used in this case study might change the results, but these results are strongly linked to the diffusion schemes, rather than to the specific values of the parameters.

5. Experimental study

A quantitative experiment must be carried out in order to evaluate the impact of the AD methods on an edge detection method, and to validate the visual assessments in Section 4. In this section we quantify the effect of preprocessing images with AD prior to extracting their edges. More specifically, we compare the performance of the Canny method for edge detection when using Gaussian Linear Filtering (GLF), used in the original proposal [47], and different AD methods.

5.1. Comparing AD and GLF

Both AD and GLF are scale-space methods [22,25], but they are based upon different paradigms. The former considers the diffusion time t, whereas the latter is ruled by the standard deviation of a Gaussian filter (σ). In order to make both paradigms comparable, we use a two-dimensional Gaussian filter G_σ, with $\sigma = t/25$, in GLF. By relating t and σ we can compare the results obtained by AD methods in the time scale-space with those obtained by GLF in the Gaussian scale-space.

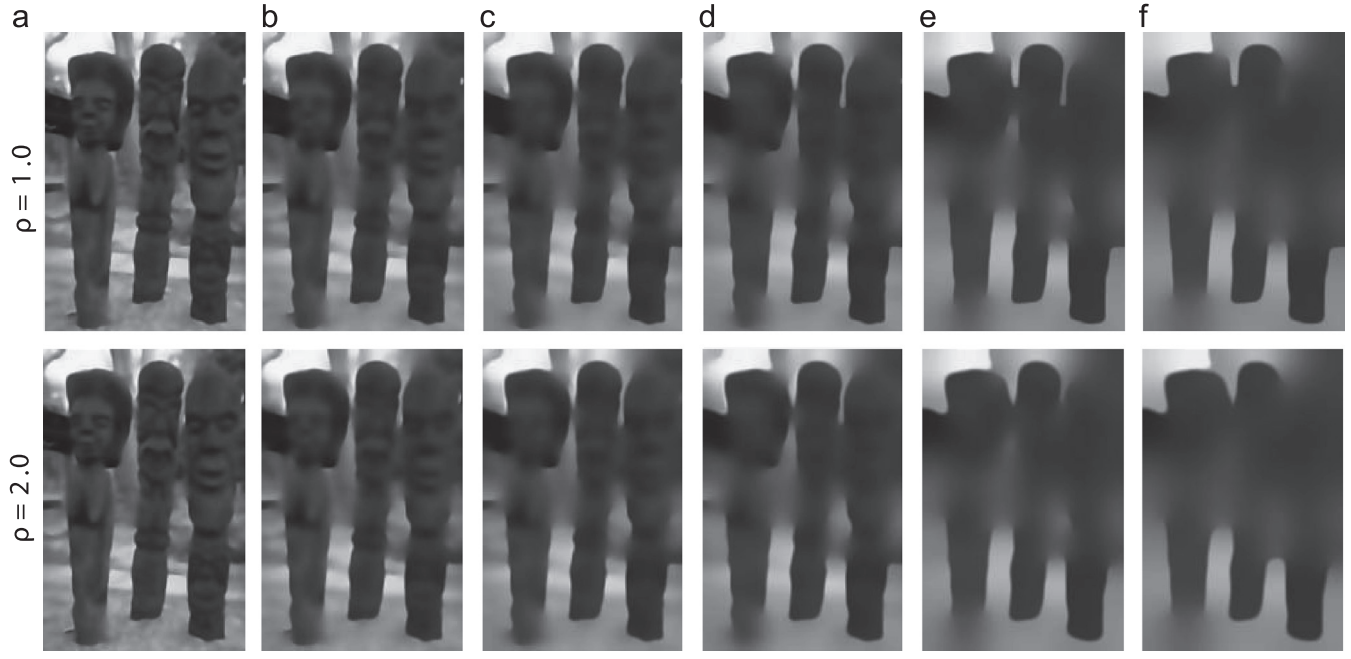

Fig. 7. Image 101085 of the BSDS applying STAD with different diffusion times t. The results of each row correspond to a different value of ρ. The parameters $\sigma_r = 0.5$, $\sigma_f = 1.0$ and $\Delta t = 0.1$ are fixed. (a) $t=10$, (b) $t=25$, (c) $t=50$, (d) $t=100$, (e) $t=300$ and (f) $t=500$.

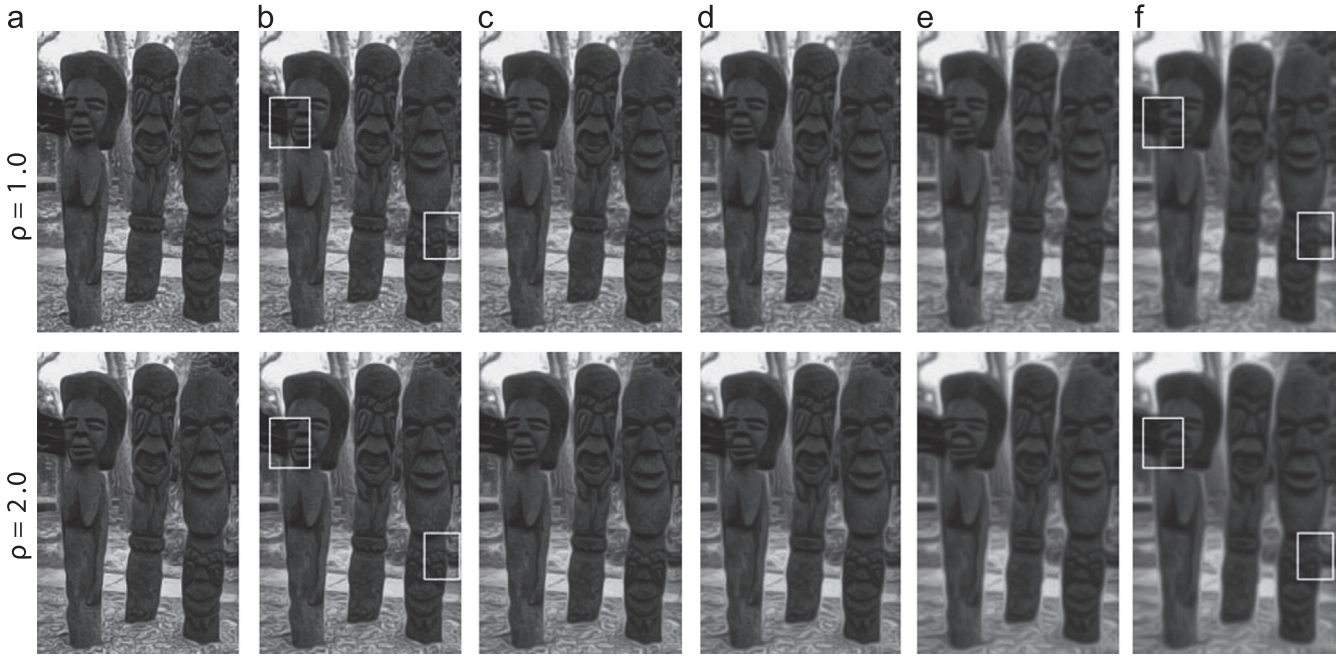

Fig. 8. Image 101085 of the BSDS applying CEAD with different diffusion times t. The results of each row correspond to a different value of ρ. The parameters $\sigma_r = 0.5$ and $\Delta t = 0.1$ are fixed. (a) $t=10$, (b) $t=25$, (c) $t=50$, (d) $t=100$ (e) $t=300$, and (f) $t=500$.

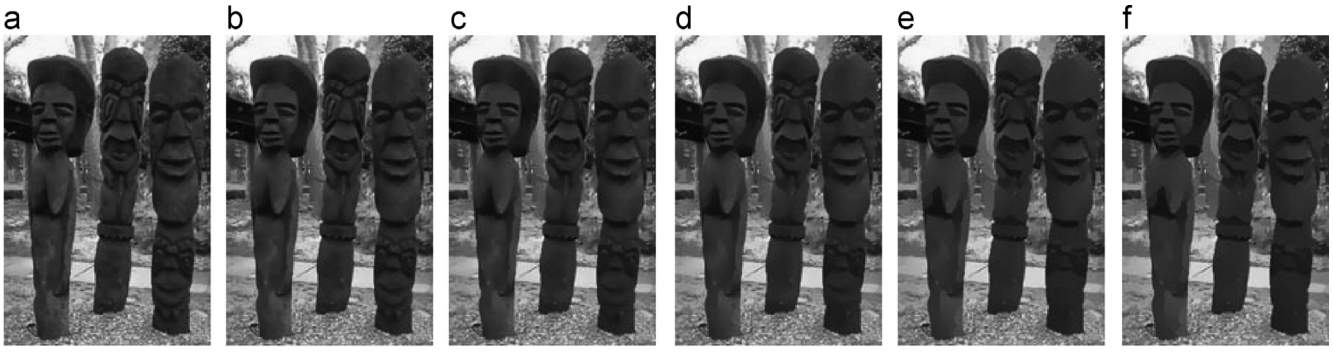

Fig. 9. Image 101085 of the BSDS applying DPAD with different diffusion times t. The parameters $\sigma_r = 0.5$ and $\Delta t = 0.1$ are fixed. (a) $t=10$, (b) $t=25$, (c) $t=50$, (d) $t=100$, (e) $t=300$ and (f) $t=500$.

5.2. Edge detection algorithm

We have selected the Canny method [3] for the experimentation. This method is commonly considered a standard, being a starting point for further developments [26,53]. Although different results might be obtained with a different edge detection method, the Canny method is not likely to bias the results, since its edge extraction operators perform well on both non-contaminated step edges (such as those by AD) and contaminated, Gaussian-filtered edges (such as those by GLF) [3,47]. Other edge detection methods have been discarded for different reasons. Those methods using information from textures [54,55] might be penalized if the image is regularized. As for the methods using training [56,57], we assume that the training after AD will probably restrict to learning *easy* or *extremely difficult* examples, since, as seen in Section 4, AD tends to either enhance or completely destroy edges. Multiscale methods [58,59] have also been discarded, since they make use of specific representations of the image in order to combine accuracy and good discrimination. Consequently, they do not demand an adaptive smoothing, such as the one performed by AD methods. In general, all the discarded methods are too similar to the Canny method, too simplistic, or significantly more complex than the Canny method, which might hinder the actual impact of the diffusion process in the improvement of the visibility of the edges.

The Canny operators are constructed with $\sigma_f = 1.0$. The binarization process consists of (a) thinning using Non-Maximum Suppression (NMS) [60] and (b) binarizing using hysteresis [3], whose thresholds are determined using the technique by Medina-Carnicer et al. [61].

5.3. Experimental dataset

We have used the 100 images in the *test* subset of the Berkeley Segmentation Dataset (BSDS) [46]. This dataset is the most popular one in the edge detection literature [62] and, since it is composed of natural images, there is no need to incorporate artificial noise. These images have a resolution of 481×321 pixels and are provided in grayscale. Each of them comes together with 5 to 10 hand-made segmentations, which are all considered in the evaluation. Since those segmentations are arranged in the shape of region boundaries, we have used them as ground truth for the quantification of the quality of the edge detection methods.

5.4. Quantification of the results

How to evaluate the performance of an edge detector is an open problem. In this work we use the approach of Martin et al. [54], which considers edge detection as a binary classification problem. Hence, it can be evaluated in terms of success and fall-out. Other alternatives, more specifically symmetric distances [63] or Baddeley's Delta Metric [64] have been discarded because they tend to overpenalize images with false detections, and hence might be prone to favor AD methods destroying a large part of the edge structures, even if they remove a significant part of the edges (see [65] for more details).

In this experimental study, we compare the output of the Canny method (candidate edge image) with the edge images generated by humans (ground truth). Each pixel in the candidate edge image is classified using the confusion matrix in Fig. 10.

When classifying the edge pixels as correct or incorrect, there are some considerations to be taken into account, due to the special conditions of the edge detection problem. More specifically, we have to consider that edges contain spatial information, and hence edge pixels should not be evaluated independently. In order to address this problem, we use a one-to-one pixel matching algorithm to map the edge pixels in the candidate edge image to those in the ground truth [66]. This matching allows for a certain spatial tolerance, so that edge pixels can be slightly displaced from their true position, yet accounting for correct classifications. The pixel-to-pixel matching is carried out using the Cost Scaling Algorithm [67], for which the maximum distance between matched pixels has been set to 14 (about 2.5% of the length of the image diagonal).

From the confusion matrix in Fig. 10 we compute the precision (PREC) and recall (REC), defined as

$$\text{PREC} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad \text{and} \quad \text{REC} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{15}$$

These measures are preferred for quantifying the performance of an edge detection method over other alternatives commonly considered for bivariate plots (e.g. ROC curves [68]) because they have good stability properties when the image size increases [69]. Moreover, they do not use TN, which is generally much larger than the other elements in the confusion matrix, and hence distorts the results. In order to evaluate the overall performance we use the *F*-measure, defined as

$$F_\alpha = \frac{\text{PREC}\cdot\text{REC}}{\alpha\text{PREC}+(1-\alpha)\text{REC}}, \tag{16}$$

where $\alpha\in[0,1]$ is a parameter weighing the contributions of PREC and REC. In this work, we set $\alpha=0.5$, so that F_α becomes the harmonic mean [70]. In this way, we evaluate three different aspects of the problem: the accuracy (using PREC), the fall-out (using REC) and the overall quality (using $F_{0.5}$).

For each image in the BSDS test set, we compare the result of the edge detection method with all of the human-made solutions. Then, the triplet (PREC, REC, $F_{0.5}$) having the greatest $F_{0.5}$ is considered as the evaluation of that image.

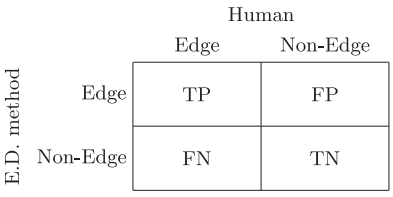

Fig. 10. Confusion matrix for the edge detection problem.

5.5. Analysis of the results

For this experiment we have used the same parameter settings used for the case study in Section 4. The results are displayed in Fig. 11. Each row is devoted to one of the evaluations (PREC, REC and $F_{0.5}$), while the results for each value of σ_r are shown in a different column. Note that PMAD and GLF are independent of the value of σ_r, since no regularization is performed prior to computing the gradients (in the former) and no adaptive behaviour is adopted (in the latter), but their results are displayed in each plot for a convenient comparison with other techniques.

- Precision: First, we note that the results for every σ_r are similar. DTAD and STAD are the best performers, mainly because PREC only penalizes FPs. As shown in Section 4, these methods are extremely destructive, and lead to an almost complete removal of spurious responses. Indeed, PREC does not reach its maximum value because some objects of the image are merged, producing new silhouettes, which are detected as edges and subsequently counted as FPs in the evaluation. PMAD does not perform so well, its PREC being consistently 0.25 worse than that of DTAD and STAD. This is explained by the staircasing effect, which produces a significant number of FPs. CEAD obtains low PREC compared to those of DTAD, STAD and PMAD. This fact was also expected given the visual experiments in Section 4, where we observed how non-meaningful structures were kept, and even reinforced. Note that the very purpose of CEAD is to emphasize significant structures, not to remove unimportant ones. We also observe that the impact of DPAD on PREC is low, the small improvement being solely due to the noisy and texturized regions it is able to regularize. All the AD methods result in an improvement of PREC as diffusion time increases, which is not the case for GLF. As expected, GLF produces a fast rise in PREC, but around $t=150$ (i.e. $\sigma=6$) it starts decreasing due to overblurring.
- Recall: The results in terms of REC are completely opposite to those regarding PREC. Indeed, the improvement of the PREC of the Canny method using AD is almost symmetrical to its loss of REC. Those AD methods performing better in terms of PREC (that is, producing a lower number of TPs) are also responsible for a greater number of FNs, i.e., the process of destroying the spurious responses comes together with the destruction of relevant edges. Even though AD aims to improve PREC (reducing the FPs) while preserving or improving REC (maintaining the TPs), this goal is not fully achieved on natural images. As shown in Section 4, the bad conditioning of natural images leads to the destruction of edges in the image, which results in a progressive decrease of REC.
- F-measure: Both PREC and REC evaluate specific facts, whereas $F_{0.5}$ provides a general quality evaluation. The most noticeable fact is that DTAD and STAD show a peak of performance (at $t\approx15$). This is exclusively due to the characteristics of $F_{0.5}$, which produces a higher result when the values to be aggregated (in this case, PREC and REC) are similar, rather than different. This is why a peak is produced when PREC$\approx$0.55 and REC$\approx$0.55. A similar peak occurs in the case of GLF. This averaging behaviour also explains why the best results are produced by PMAD, since it represents an equilibrated trade-off between the most destructive methods, with high PREC and low REC (STAD, DTAD), and those having a more conservative behaviour, which provide higher REC evaluations (CEAD, DPAD).

In general, we observe that the quantitative results are consistent with the conclusions drawn in the visual case study in Section 4. In addition, the practical invariance with respect to values of σ_r and ρ is remarkable. This is due to the fact that different images (or even different regions of the same image)

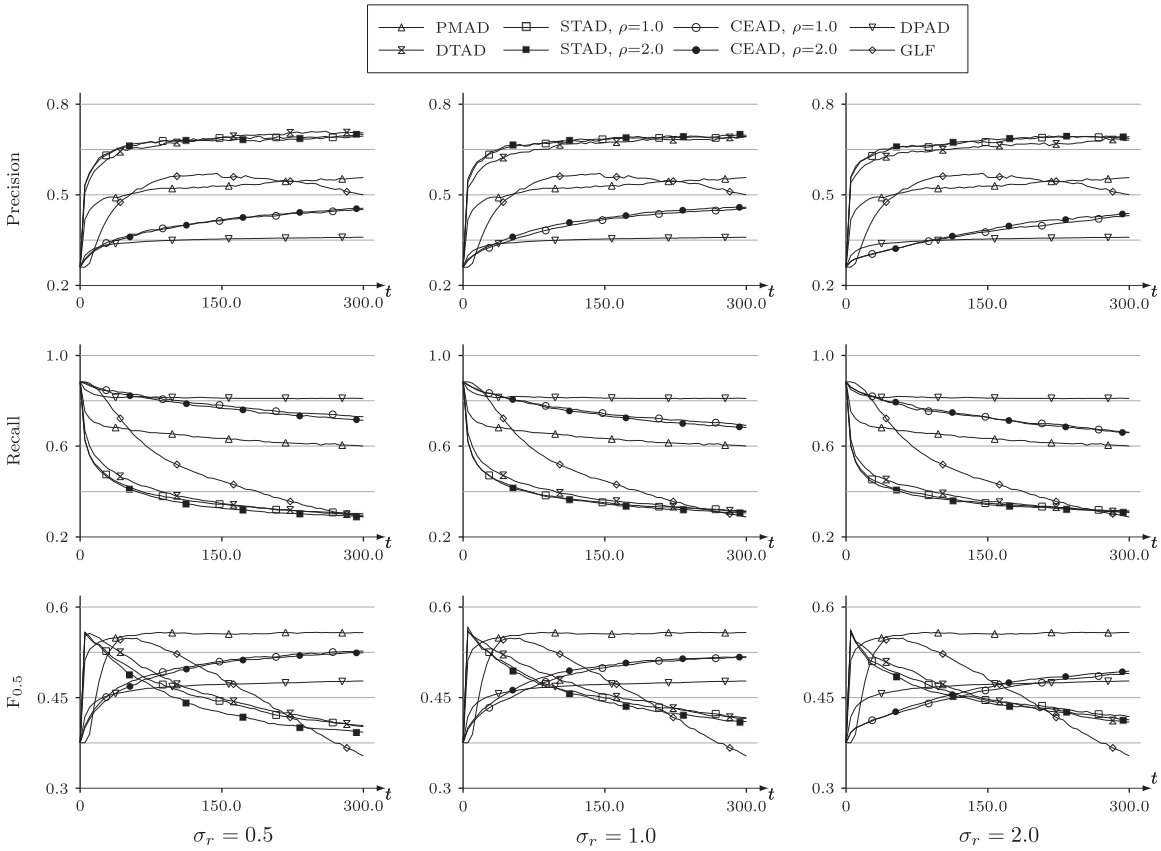

Fig. 11. Results obtained by different AD methods on the BSDS test set using three different values of σ and $\Delta t = 0.1$. The results are quantified using PREC, REC and $F_{0.5}$. In the comparison we also include Gaussian Linear Filtering with $\sigma = t/25$.

have different characteristics and demand different settings. There is no single solution for the scale determination problem [26,43]. As shown in Fig. 1, different objects in the image demand different values of σ_r (or ρ). Therefore, the use of different values of σ_r leads to different visual results but, from a quantitative point of view, the improvements and degradations are compensated. The same occurs for the parameter ρ. Since there is no optimal setting of the parameter value, any value within a reasonable range produces better results for some images and worse results for others, leading to similar average evaluations.

6. Conclusions

In this work we have reviewed some representative AD methods in the literature, paying special attention to the way in which they manage local information. We have used a visual example to illustrate the results they produce. We have also performed a quantitative evaluation of their impact on the results of an edge detection method. Our interest is in how much the results can be improved, compared to those by GLF, as well as how robust that improvement is against changes in the parameters of the AD methods.

Regarding the potential improvement induced by AD, our conclusion is that the results of an edge detector can be improved by using AD instead of GLF. This can lead to large potential improvements in terms of precision and recall, as reflected in Fig. 11. However, we find that the increase of the quality in terms of PREC is related to the decrease in terms of REC, and vice versa. In that sense, the AD methods do not reach their goal of preserving edges while removing spurious artifacts of the image. As a consequence, the improvement of the total quality (in this case, $F_{0.5}$) is rather marginal. There are two different annotations to be made to this conclusion:

(a) The performance of the AD methods is limited by the conditions of the natural images, which lead to undesired results even if the diffusion occurs according to the expected behaviour. We observed in our case study how diffusion methods can damage important structures even if the diffusion processes behave as expected.
(b) Different AD methods lead to different effects on PREC and REC. This indicates that each of them has a variable suitability, depending on facts such as the relevance of the false (mis-) detections for that specific task or the presence of a large number of false detections due to the conditions of the images. However, the selection of the most appropriate AD method is tightly bounded to the specific application.

Regarding the stability of the results w.r.t. changes in the parameters, we have observed that even though the parameter settings have a significant impact on the visual results, their highs and lows compensate so that no configuration seems to be clearly better than the others from a quantitative point of view. As for the diffusion time, we have observed that AD methods hardly ever reach completely stable states [26], although most of the changes the image undergoes happen in the first instants of diffusion.

As a general conclusion, we have that AD partially reaches its goals. Its results are visually interesting and have a noteworthy impact on the results of an edge detection method. However, although significantly attenuated, it suffers from the same problems GLF does: the destruction of non-meaningful elements is coupled to the destruction of meaningful edges.

Conflict of interest

None declared.

Acknowledgments

This work was partially funded by the Spanish Ministry of Science, project TIN2010-15055, and by the Research Services of the Universidad Publica de Navarra.

References

[1] D. Marr, E. Hildreth, Theory of edge detection, Proceedings of the Royal Society of London 207 (1167) (1980) 187–217.
[2] R.M. Haralick, Digital step edges from zero crossing of second directional derivatives, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1) (1984) 58–68.
[3] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6) (1986) 679–698.
[4] M. Basu, Gaussian-based edge-detection methods—A survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 32 (3) (2002) 252–260.
[5] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7) (1990) 629–639.
[6] J. Monteil, A. Beghdadi, A new interpretation and improvement of the nonlinear anisotropic diffusion for image enhancement, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (9) (1999) 940–946.
[7] M. Black, G. Sapiro, D. Marimont, D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing 7 (3) (1998) 421–432.
[8] J. Song, H.R. Tizhoosh, Fuzzy anisotropic diffusion based on edge detection, Journal of Intelligent and Fuzzy Systems 17 (5) (2006) 431–442.
[9] J. Lee, X. Geets, V. Gregoire, A. Bol, Edge-preserving filtering of images with low photon counts, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (6) (2008) 1014–1027.
[10] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, H.-P. Seidel, Image compression with anisotropic diffusion, Journal of Mathematical Imaging and Vision 31 (2008) 255–269.
[11] M. Mainberger, A. Bruhn, J. Weickert, S. Forchhammer, Edge-based compression of cartoon-like images with homogeneous diffusion, Pattern Recognition 44 (9) (2011) 1859–1873.
[12] C. Schmaltz, J. Weickert, A. Bruhn, Beating the quality of JPEG 2000 with anisotropic diffusion, in: Pattern Recognition, Lecture Notes in Computer Science, vol. 5748, Springer, Berlin/Heidelberg, 2009, pp. 452–461.
[13] P. Meer, B. Georgescu, Edge detection with embedded confidence, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 1351–1365.
[14] S. Yi, D. Labate, G. Easley, H. Krim, A shearlet approach to edge analysis and detection, IEEE Transactions on Image Processing 18 (5) (2009) 929–941.
[15] O. Laligant, F. Truchetet, A nonlinear derivative scheme applied to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2) (2010) 242–257.
[16] S.M. Smith, J.M. Brady, SUSAN – a new approach to low level image processing, International Journal of Computer Vision 23 (1997) 45–78.
[17] S. Weeratunga, C. Kamath, A comparison of PDE-based non-linear anistropic diffusion techniques for image denoising, in: E.R. Dougherty, J.T. Astola, K.O. Egiazarian (Eds.), Image Processing: Algorithms and Systems II, vol. 5014 of Proceedings of the SPIE, 2003, pp. 201–212.
[18] Y. Yu, S. Acton, Speckle reducing anisotropic diffusion, IEEE Transactions on Image Processing 11 (11) (2002) 1260–1270.
[19] T. Peli, D. Malah, A study of edge detection algorithms, Computer Graphics and Image Processing 20 (1) (1982) 1–21.
[20] G. Papari, N. Petkov, Edge and line oriented contour detection: state of the art, Image and Vision Computing 29 (2–3) (2011) 79–103.
[21] J. Weickert, B. ter Haar Romeny, M. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, IEEE Transactions on Image Processing 7 (3) (1998) 398–410.
[22] J. Weickert, Anisotropic Diffusion in Image Processing, ECMI Series, Teubner-Verlag, 1998.
[23] G. Cottet, L. Germain, Image processing through reaction combined with nonlinear diffusion, Mathematics of Computation 61 (204) (1993) 659–673.
[24] S. Aja-Fernández, C. Alberola-López, On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering, IEEE Transactions on Image Processing 15 (9) (2006) 2694–2701.
[25] A.P. Witkin, Scale-space filtering, in: Proceedings of the International Joint Conference on Artificial Intelligence, vol. 2, Karlsruhe, 1983, pp. 1019–1022.
[26] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 117–156.
[27] P. Saint-Marc, J.-S. Chen, G. Medioni, Adaptive smoothing: a general tool for early vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (6) (1991) 514–529.
[28] A. Goshtasby, M. Satter, An adaptive window mechanism for image smoothing, Computer Vision and Image Understanding 111 (2) (2008) 155–169.
[29] F. Catté, P.-L. Lions, J.-M. Morel, T. Coll, Image selective smoothing and edge detection by nonlinear diffusion, SIAM Journal on Numerical Analysis 29 (1992) 182–193.
[30] M. Nitzberg, T. Shiota, Nonlinear image filtering with edge and corner enhancement, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (8) (1992) 826–833.
[31] R.T. Whitaker, S.M. Pizer, A multi-scale approach to nonuniform diffusion, CVGIP: Image Understanding 57 (1) (1993) 99–110.
[32] Y.-L. You, W. Xu, A. Tannenbaum, M. Kaveh, Behavioral analysis of anisotropic diffusion in image processing, IEEE Transactions on Image Processing 5 (11) (1996) 1539–1553.
[33] J. Weickert, Theoretical foundations of anisotropic diffusion in image processing, Computing 11 (1996) 221–236.
[34] J. Weickert, Coherence-enhancing diffusion filtering, International Journal of Computer Vision 31 (2–3) (1999) 111–127.
[35] S. Acton, Multigrid anisotropic diffusion, IEEE Transactions on Image Processing 7 (3) (1998) 280–291.
[36] S. Keeling, R. Stollberger, Nonlinear anisotropic diffusion filtering for multiscale edge enhancement, Inverse Problems 18 (1) (2002) 175–190.
[37] N. Sochen, R. Kimmel, A. Bruckstein, Diffusions and confusions in signal and image processing, Journal of Mathematical Imaging and Vision 14 (2001) 195–209.
[38] S. Aja, C. Alberola, A. Ruiz, Fuzzy anisotropic diffusion for speckle filtering, in: Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 2001, pp. 1261–1264.
[39] D. Tschumperle, R. Deriche, Vector-valued image regularization with PDEs: a common framework for different applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 506–517.
[40] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Proceedings of the International Conference on Computer Vision, 1998, pp. 838–846.
[41] D. Barash, A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 844–847.
[42] M. Alemán-Flores, L. Álvarez, V. Caselles, Texture-oriented anisotropic filtering and geodesic active contours in breast tumor ultrasound segmentation, Journal of Mathematical Imaging and Vision 28 (2007) 81–97.
[43] V. Torre, T. Poggio, On edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (2) (1984) 147–163.
[44] R. Deriche, Using Canny's criteria to derive a recursively implemented optimal edge detector, International Journal of Computer Vision (1987) 167–187.
[45] J. Babaud, A.P. Witkin, M. Baudin, R.O. Duda, Uniqueness of the Gaussian kernel for scale-space filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1) (1986) 26–33.
[46] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the Eighth International Conference on Computer Vision, vol. 2, 2001, pp. 416–423.
[47] J. Canny, Finding Edges and Lines in Images, Technical Report, Massachussets Institute of Technology, Cambridge, MA, USA, 1983.
[48] K. Abd-Elmoniem, A.-B. Youssef, Y. Kadah, Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion, IEEE Transactions on Biomedical Engineering 49 (9) (2002) 997–1014.
[49] K. Krissian, C.-F. Westin, R. Kikinis, K. Vosburgh, Oriented speckle reducing anisotropic diffusion, IEEE Transactions on Image Processing 16 (5) (2007) 1412–1424.
[50] J. Weickert, Nonlinear diffusion scale-spaces: from the continuous to the discrete setting, in: M.-O. Berger, R. Deriche, I. Herlin, J. Jaffré, J.-M. Morel (Eds.), ICAOS '96, Lecture Notes in Control and Information Sciences, vol. 219, Springer, Berlin, Heidelberg, 1996, pp. 111–118.
[51] J. Weickert, Recursive separable schemes for nonlinear diffusion filters, in: B. Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, Lecture Notes in Computer Science, vol. 1252, Springer, Berlin, Heidelberg, 1997, pp. 260–271.
[52] P.L. Rosin, Unimodal thresholding, Pattern Recognition 34 (11) (2001) 2083–2096.
[53] P.-L. Shui, W.-C. Zhang, Noise-robust edge detector combining isotropic and anisotropic Gaussian kernels, Pattern Recognition 45 (2) (2012) 806–820.
[54] D. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 530–549.
[55] J. Malik, S. Belongie, T. Leung, J. Shi, Contour and texture analysis for image segmentation, International Journal of Computer Vision 43 (2001) 7–27.
[56] F. Russo, Edge detection in noisy images using fuzzy reasoning, in: Instrumentation and Measurement Technology Conference, vol. 1, 1998, pp. 369–372.
[57] J. Bezdek, R. Chandrasekhar, Y. Attikouzel, A geometric approach to edge detection, IEEE Transactions on Fuzzy Systems 6 (1) (1998) 52–75.
[58] G. Papari, P. Campisi, N. Petkov, A. Neri, A biologically motivated multiresolution approach to contour detection, EURASIP Journal on Advances in Signal Processing (2007), < http://asp.eurasipjournals.com/content/2007/1/071828 > .
[59] C. Lopez-Molina, B. De Baets, H. Bustince, J. Sanz, E. Barrenechea, Multiscale edge detection based on Gaussian smoothing and edge tracking, Knowledge-Based Systems 44 (2013) 101–111.

[60] A. Rosenfeld, M. Thurston, Edge and curve detection for visual scene analysis, IEEE Transactions on Computers 20 (5) (1971) 562–569.
[61] R. Medina-Carnicer, F. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, On candidates selection for hysteresis thresholds in edge detection, Pattern Recognition 42 (7) (2009) 1284–1296.
[62] X. Hou, A. Yuille, C. Koch, A meta-theory of boundary detection benchmarks, in: Proceedings of the NIPS Workshop on Human Computation for Science and Computational Sustainability, 2012.
[63] T. Heimann, B. van Ginneken, M. Styner, et al., Comparison and evaluation of methods for liver segmentation from CT datasets, IEEE Transactions on Medical Imaging 28 (8) (2009) 1251–1265.
[64] A.J. Baddeley, An error metric for binary images, in: W. Förstner, S. Ruwiedel (Eds.), Robust Computer Vision: Quality of Vision Algorithms, Wichmann Verlag, Karlsruhe, 1922, pp. 59–78.
[65] C. Lopez-Molina, B. De Baets, M. Galar, H. Bustince, A generalization of the Perona–Malik anisotropic diffusion method using restricted dissimilarity functions, International Journal of Computational Intelligent Systems 6 (1) (2013) 14–28.
[66] G. Liu, R.M. Haralick, Optimal matching problem in detection and recognition performance evaluation, Pattern Recognition 35 (10) (2002) 2125–2139.
[67] A.V. Goldberg, R. Kennedy, An efficient cost scaling algorithm for the assignment problem, Mathematical Programming 71 (1995) 153–177.
[68] W. Waegeman, B. De Baets, L. Boullart, ROC analysis in ordinal regression learning, Pattern Recognition Letters 29 (1) (2008) 1–9.
[69] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2011) 898–916.
[70] G. Beliakov, A. Pradera, T. Calvo, in: Aggregation Functions: A Guide for Practitioners, Studies in Fuzziness and Soft Computing, vol. 221, Springer, 2007.

Carlos Lopez-Molina (1985) received the Ph.D. degree in computer sciences from the Public University of Navarra, Pamplona, Spain, in 2012. He currently holds a full-time research position with the Department of Automatics and Computation, Public University of Navarra, where he is involved in teaching scientific computing. His research interests include edge detection, image processing, aggregation operators, and fuzzy logic applications.

Mikel Galar received the M.Sc. and Ph.D. degrees in Computer Science in 2009 and 2012, both from the Public University of Navarra, Pamplona, Spain. He is currently a teaching assistant in the Department of Automatics and Computation at the Public University of Navarra. His research interests are data-mining, classification, multi-classification, ensemble learning, evolutionary algorithms, fuzzy systems and image processing.

Humberto Bustince (M'08) received the Ph.D. degree in mathematics from the Public University of Navarra, Pamplona, Spain, in 1994. He is a Full Professor with the Department of Automatics and Computation, Public University of Navarra, Spain. His research interests include fuzzy logic theory, extensions of fuzzy sets (type-2 fuzzy sets, interval-valued fuzzy sets, Atanassov intuitionistic fuzzy sets), fuzzy measures, aggregation functions, and fuzzy techniques for image processing. He is the author of more than 70 published original articles and is involved in teaching artificial intelligence to students of computer science.

Bernard De Baets (1966) holds an M.Sc. in Maths (1988), a Postgraduate degree in Knowledge Technology (1991) and a Ph.D. in Maths (1995), all summa cum laude from Ghent University (Belgium) and is a Government of Canada Award holder (1988). He is a Full Professor in Applied Maths (1999) at Ghent University, where he is leading KERMIT, the research unit Knowledge-Based Systems. He is an Honorary Professor of Budapest Tech (2006) and an IFSA Fellow (2011). His publications comprise more than 250 papers in international journals and about 50 book chapters. He serves on the Editorial Boards of various international journals, in particular as co-editor-in-chief of Fuzzy Sets and Systems. B. De Baets coordinates EUROFUSE, the EURO Working Group on Fuzzy Sets, and is a member of the Board of Directors of EUSFLAT and of the Administrative Board of the Belgian OR Society.