Deterministic Edge-Preserving Regularization in Computed Imaging

Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud, *Member, IEEE*

***Abstract*—Many image processing problems are *ill posed* and must be regularized. Usually, a roughness penalty is imposed on the solution. The difficulty is to avoid the smoothing of edges, which are very important attributes of the image. In this paper, we first give conditions for the design of such an *edge-preserving regularization*. Under these conditions, we show that it is possible to introduce an auxiliary variable whose role is twofold. First, it marks the discontinuities and ensures their preservation from smoothing. Second, it makes the criterion *half-quadratic*. The optimization is then easier. We propose a deterministic strategy, based on alternate minimizations on the image and the auxiliary variable. This leads to the definition of an original reconstruction algorithm, called ARTUR. Some theoretical properties of ARTUR are discussed. Experimental results illustrate the behavior of the algorithm. These results are shown in the field of tomography, but this method can be applied in a large number of applications in image processing.**

I. Introduction

In COMPUTED imaging, reconstructing an image f from data p is often an ill-posed problem in the sense of Hadamard. Knowledge of the direct model is not always sufficient to determine a satisfying solution, and it is necessary to *regularize* the solution by imposing an *a priori* constraint. Mathematically, this constraint is often expressed through a regularization function—which is also called a *potential function* in the Markov random field approach [15]. A simple and well-known regularization supposes that images are globally smooth, and enforces a roughness penalty on the solution. A quadratic potential function yields oversmooth solutions. A more realistic image model assumes that images are made of smooth regions, separated by sharp edges [15]. This is called *edge-preserving regularization* and requires a nonquadratic potential function.

Edge-preserving regularization has provided an abundant literature in the last decade, but curiously, to our knowledge there is no unified theory about the design of potential functions for this purpose. For example, the question: "what properties must a potential function satisfy to ensure the preservation of edges?" has different and sometimes contradictory answers. For Geman and Reynolds, one of the two important attributes of an edge-preserving potential function is its finite asymptotic behavior [17], while other authors advocate convex functions [6], [20], [26]. In our first contribution of this paper, we try to unify these different approaches by proposing three *conditions for edge preservation*. These define a class of potential functions which yield edge-preserving regularization.

An important characteristic of edge-preserving regularization is that the computations involve the minimization of possibly nonconvex energy functionals. In many applications computation time is critical, so a deterministic strategy is preferable. Also, one has to face the problem of minimizing nonquadratic energy functionals or, equivalently, solving nonlinear simultaneous equations. In our second contribution of this paper, we show that, when the conditions for edge preservation are satisfied, it is possible to transform the nonquadratic energy into an *augmented energy* by introducing an auxiliary variable, b, whose role is twofold. First, b marks the location of discontinuities, and thus takes part in their preservation. Second, b makes the augmented energy functional become *half-quadratic*, i.e., quadratic with respect to the image variable when b is fixed. We also show that the augmented functional is convex with respect to b when the image variable is fixed, and we give an exact expression for the minimum b.

Using the auxiliary variable permits us to *linearize* the problem and to derive a *deterministic* algorithm based on alternate minimizations on the image variable and the auxiliary variable. This relaxation algorithm, called ARTUR, is based on the general principle of minimizing a sequence of energy functionals. Well-known deterministic algorithms, such as graduated nonconvexity (GNC) [3] or mean-field annealing (MFA) [14], use the same principle. However, in contrast to these methods, ARTUR can be applied to many energies having an edge-preserving potential function and to many types of inverse problems (tomography, restoration, motion estimation, stereovision, etc.). In our third contribution of this paper, we show experimental results in tomography, with both synthetic and real data, to illustrate the way the algorithm works and the behavior of different potential functions.

This paper is organized as follows. The conditions for edge-preservation are defined and discussed in Section II. In Section III, we discuss edge-preserving regularization and the principle of half-quadratic regularization. Sections IV and V are, respectively, devoted to the theoretical and experimental presentation of ARTUR. Section VI concludes the paper.

II. Edge-Preserving Regularization

In this paper, the image and auxiliary variables will be either considered as $N \times N$ two-dimensional (2-D) fields

Manuscript received March 4, 1994; revised April 16, 1996.

P. Charbonnier, L. Blanc-Féraud, and M. Barlaud are with Laboratoire I3S, CNRS-UNSA, GDR Traitement du Signal et Images, Université de Nice-Sophia Antipolis, 06560 Valbonne, France (e-mail: cp@alto.unice.fr)..

G. Aubert is with the Laboratoire J. A. Dieudonné, URA 168 CNRS, Université de Nice-Sophia Antipolis, 6108 Nice Cedex 2, France (e-mail: aubert@orque.unice.fr).

Publisher Item Identifier S 1057-7149(97)00485-5.

indexed by i (row number) and j (column number), or as $N^2 \times 1$ lexicographically ordered (column) vectors indexed by k. The index transformation between the two representations is $k = i \times N + j$.

A. Regularized Image Reconstruction

In many computed imaging applications, the observed data p can be related to the original image, f, through a linear model of the form

$$p = Rf + \eta \tag{1}$$

where η is assumed to be white Gaussian noise, and R is a linear operator. In image restoration, for example, R is block-Toeplitz and represents the point spread function (PSF) of the imaging system [23].

In computed tomography (CT), R models the Radon transform [23]. The reader might object that a white Gaussian noise model is not realistic in CT, because it is well known that the photon counts have a Poisson statistic. Though all that is developed in this paper can be adapted to the case of the Poisson model, we use the Gaussian approximation because it yields additive algorithms while the Poisson model leads to multiplicative—and hence slower—algorithms. However, for a sufficient number of counts, the Poisson distribution can be successfully approximated by a Gaussian one (as in [13] for example). Theoretically, our noise model should include a covariance matrix. We consider *white* noise for the sake of simplicity and because we have experimentally found that our model yields good results even for low photon dosages [25]. Anyway, considering a general covariance matrix would make no theoretical or practical difficulty.

For some other applications (as stereovision and motion estimation [4], [22]) the model is nonlinear, but approximations can be made to linearize it, and all that will be developed hereafter can be applied in these cases too.

The problem of regularized image reconstruction has been widely investigated in the past (see for example [12] for an overview), with two main approaches. On one hand, some authors [27], [29], [32] consider the problem from a deterministic point of view. On the other hand, other authors [2], [6], [7], [15], [24], [28] use a stochastic approach. We consider a framework either called penalized least-squares or maximum *a posteriori* (MAP) estimation, in which the estimated image $\hat{f}$ is given by

$$\hat{f} = \arg\min_f (J(f)) \tag{2}$$

where J is the sum of a term which measures the faithfulness of the estimate to the data and a regularization term

$$J(f) = J_1(f) + \lambda^2 J_2(f). \tag{3}$$

In our case, the data term takes the following form:

$$J_1(f) = \|p - Rf\|^2. \tag{4}$$

The regularization term is defined as a sum of potentials which are, in general, functions of a derivative of the image. The order of the derivative depends on the kind of image that is sought. For example, the use of second-order derivatives promotes the formation of piecewise planar areas in the solution [17]. In our case, though, the theory we develop is also valid for higher-order derivatives, we only use a simple model in which the image is supposed to be piecewise constant. Therefore, we consider first-order differences between pixels belonging to the two-neighbor cliques of a second-order model [15] (see Fig. 1). In order to simplify our presentation, we "forget" about diagonal differences for the moment. Their case will be discussed in Section IV-D. This leads to the following expression for the regularization term

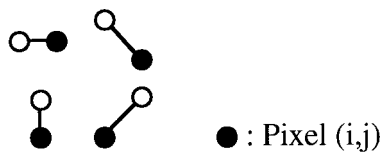


Fig. 1. Cliques of the model at site (i, j).

$$J_2(f) = \sum_k \varphi[(D_x f)_k] + \sum_k \varphi[(D_y f)_k] \tag{5}$$

where

$$(D_x f)_{i,j} = (f_{i,j+1} - f_{i,j})/\delta$$

and

$$(D_y f)_{i,j} = (f_{i+1,j} - f_{i,j})/\delta. \tag{6}$$

λ^2 in (3) and δ in (6) are the parameters of the model. The regularization coefficient, λ^2, balances the effects of the data term and the *a priori* term. The second parameter, δ, is a scaling parameter which tunes the value of the gradient above which a discontinuity is detected.

In the Bayesian framework, φ in (5) is known as the *potential function*. This function assigns a cost to every value of the image gradient, and thus should have some obvious properties. First, it seems natural to give positive values to the potential. We will therefore require that $\varphi(t) \geq 0 \forall t$, with $\varphi(0) = 0$ for practical reasons. Second, in designing a potential function, it is natural to assume that φ is an increasing function for $t \geq 0$. Also, it is necessary to give the same importance to gradients of equal values but opposite signs. Thus, φ is assumed to be an even function. We can then limit our study to positive values of the gradient.

In order to avoid introducing instability into the reconstruction process, differentiability is desirable for φ. We will focus on continuously differentiable potential functions. Note that this excludes functions like $|t|$ [2], which are not differentiable at 0, or like the truncated quadratic [3].

These general assumptions are summarized in 12(a)–(d). In Section II-B, we discuss desirable properties of the potential functions for edge preservation.

B. Edge-Preserving Regularization

A priori information imposed on the solution is expressed via the potential function. In this section, we try to answer the following question: "What properties should the potential function satisfy to define an edge-preserving regularization?" Curiously, this question has several different answers in the

$$\begin{pmatrix} 0 & \lambda^{\mathcal{N}} & 0 \\ \lambda^{\mathcal{W}} & -\Sigma & \lambda^{\mathcal{E}} \\ 0 & \lambda^{\mathcal{S}} & 0 \end{pmatrix}$$

$\lambda^{\mathcal{E}} = \dfrac{\varphi'(f_{i,j+1} - f_{i,j})}{2(f_{i,j+1} - f_{i,j})}$	$\lambda^{\mathcal{W}} = \dfrac{\varphi'(f_{i,j} - f_{i,j-1})}{2(f_{i,j} - f_{i,j-1})}$
$\lambda^{\mathcal{S}} = \dfrac{\varphi'(f_{i+1,j} - f_{i,j})}{2(f_{i+1,j} - f_{i,j})}$	$\lambda^{\mathcal{N}} = \dfrac{\varphi'(f_{i,j} - f_{i-1,j})}{2(f_{i,j} - f_{i-1,j})}$

$$\Sigma = \lambda^{\mathcal{N}} + \lambda^{\mathcal{S}} + \lambda^{\mathcal{E}} + \lambda^{\mathcal{W}}$$

Fig. 2. Coefficients of the weighted Laplacian around pixel (i, j) (first-order neighborhood).

literature. For example, Geman and Reynolds [17] advocate functions having a finite asymptotic behavior. Other authors, as Green [20], Bouman and Sauer [6], Schultz and Stevenson [31], and Lange [26] prefer using convex potentials in order to ensure uniqueness of the solution. Lastly, other authors, as Hebert and Leahy [21] propose a compromise between both approaches. Having a view to unify these approaches, we propose a local heuristic study of the first-order necessary conditions associated with the minimization of the energy in (3). Suppose that J has a minimum in f, then we have necessarily

$$\frac{1}{2}J'(f) = 0 \tag{7}$$

where J' is the derivative of J. A simple calculation (details are given in Appendix A) shows that, (7) can be written as

$$R^t Rf - R^t p - \lambda^2 \Delta_{\text{pond}} f = 0 \tag{8}$$

where Δ_{pond} is a matrix that represents a weighted discrete approximation of the Laplacian operator. The matrix-vector multiplication $\Delta_{\text{pond}} f$ is equivalent to a nonstationary filtering of f by a 3×3 weighted Laplacian filter, which is shown in Fig. 2. The weights of the Laplacian are given by the function $\varphi'(t)/2t$, which we call the *weighting function*. Note that since φ is even, the weighting function is even.

Now, let us consider the case of a homogeneous area of the image: All gradients around pixel (i, j) are close to zero. Suppose that the weighting function is such that

$$\frac{\varphi'(t)}{2t} \underset{t \to 0}{\longrightarrow} M < +\infty \tag{9}$$

then, all weights around pixel (i, j) are approximately equal to M and the weighted Laplacian behaves as the usual Laplacian (see Fig. 3(a)). The necessary conditions then locally reduce to the usual normal equations associated with Tikhonov regularization

$$R^t Rf - R^t p - \lambda^2 M \Delta f = 0 \tag{10}$$

where Δf is the usual discrete Laplacian of f. In other words, there is diffusion (i.e., smoothing) all around pixel (i, j).

We define the value of the weighting function at 0 to be equal to M, with M finite. Most of the functions that can be found in the literature satisfy this condition. This is not the case when $\varphi(t) = |t|^\alpha (1 \leq \alpha < 2)$ [6]. Therefore, using these functions would create problems at 0. First, writing (10) would make no sense. Second, an infinite value of M would involve numerical problems using the algorithm we propose in Section IV.

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & -3 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Fig. 3. Coefficients of the weighted Laplacian around pixel (i, j) in a homogeneous area (a) and in the case of a discontinuity (b) (first-order neighborhood).

Now, let us suppose that there is a discontinuity in the neighborhood of pixel (i, j), for example between pixel $(i, j-1)$ and pixel (i, j). Then, all finite differences around pixel (i, j) are small, except $f_{i,j-1} - f_{i,j}$. Suppose that the weighting function is such that

$$\frac{\varphi'(t)}{2t} \underset{t \to +\infty}{\longrightarrow} 0. \tag{11}$$

Then, the corresponding weight of the Laplacian vanishes (see Fig. 3(b)) and there is no smoothing in this direction.

Lastly, we suppose that $\varphi'(t)/2t$ is continuous, because we do not want a small variation of the gradient to produce a large change in the value of the weight. Otherwise, this might produce instabilities in the presence of noise. Also, it seems natural that there should be a one-to-one correspondence between values of the gradient and values of the weight. Therefore the weighting function must be strictly monotonous on $[0, +\infty)$.

All the conditions that we impose on φ are summarized in (12).

General: Basic assumptions

a) $\varphi(t) \geq 0 \ \forall t$ with $\varphi(0) = 0$.
b) $\varphi(t) = \varphi(-t)$.
c) φ continuously differentiable.
d) $\varphi'(t) \geq 0 \ \forall t \geq 0$.

General: Edge preservation

e) $\varphi'(t)/2t$ continuous and strictly decreasing on $[0, +\infty)$.
f) $\lim_{t \to +\infty} \dfrac{\varphi'(t)}{2t} = 0$.
g) $\lim_{t \to 0^+} \dfrac{\varphi'(t)}{2t} = M, \ 0 < M < +\infty$.

Algorithm: Convergence proof

h) $\varphi'''(0) = 0$.
i) $\varphi^{(4)}(0)$ exists. (12)

The basic assumptions a) to d) define the limits of our study. Conditions e), f), and g) are the three *conditions for edge-preservation*. Additional requirements on φ, h) and i), are technical hypothesis for the convergence proofs of the algorithm we propose in Section IV.

Note that in the above study, we have considered that small gradients must be smoothed, while large gradients must be preserved. Hence, we have implicitly made the following assumption: A large value of the gradient corresponds to an edge while a small value of the gradient is an effect of noise. This assumption is not necessarily satisfied in practice: small values of the gradient may as well correspond to an actual low amplitude discontinuity in the image. On the other hand, large gradients may be due to noise. Undoubtedly, this is a limitation of the model: Especially in presence of strong noise, it may not

TABLE I
FOUR EDGE-PRESERVING POTENTIAL FUNCTIONS AND THEIR ASSOCIATED WEIGHTING FUNCTIONS

Potential function	Expression of $\varphi(t)$	Expression of $\varphi'(t)/2t$	Ref.
φ_{GM}	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	[16]
φ_{HL}	$\log(1+t^2)$	$\frac{1}{1+t^2}$	[21]
φ_{HS}	$2\sqrt{1+t^2}-2$	$\frac{1}{\sqrt{1+t^2}}$	[10] [11]
φ_{GR}	$2\log[\cosh(t)]$	$\begin{cases} 1 & t=0 \\ \tanh(t)/t & t\neq 0 \end{cases}$	[20]

always be possible to discriminate real features of the image from the effects of noise.

We would like to stress that conditions (12) are satisfied by many of the edge-preserving potential functions proposed in the literature. In Table I we give four examples of well-known potential functions and their corresponding weighting functions. The potential functions have been normalized in order to have $M = 1$ for all the weighting functions in Table I. Note that even if the potential functions have different behaviors at the infinity (some have an horizontal asymptote, as φ_{GM}, others do not), all the weighting functions in Table I satisfy (12). Contrarily to what was suggested in [17], it appears that the existence of an horizontal asymptote is not necessary to ensure edge-preservation. Moreover, it shows that edge-preservation can be achieved even by some *convex* potentials, as φ_{HS} or φ_{GR}. This is an important result since using convex potentials generally make the minimization problem well posed [6]. Furthermore, these two functions are continuously differentiable and satisfy (12 g), so they do not involve numerical difficulties as when using $\varphi(t) = |t|^\alpha$ $(1 \leq \alpha < 2)$. Note that we propose a new potential function, φ_{HS}. It has the same behavior as φ_{GR}, but its associated weighting function is simpler than that of φ_{GR}.

III. HALF-QUADRATIC REGULARIZATION

Even in the convex case, minimizing the MAP criterion is a difficult task because the necessary conditions (8) are nonlinear. In order to simplify the minimization task, we propose to use *half-quadratic regularization.*

A. Principle of Half-Quadratic Regularization

The expression "half-quadratic regularization" was defined by Geman and Yang in 1993 [18]. To quote these authors, "The basic idea is to introduce a new objective function which, although defined over an extended domain, has the same minimum in f as J and can be manipulated with linear algebraic methods." The principle is to introduce a couple of $N \times N$ auxiliary variables, $b = (b_x, b_y)$,[1] in order to make the manipulation of the MAP criterion easier.

[1] For clarity, we will use hereafter b instead of (b_x, b_y) whenever possible.

When φ is edge-preserving in the sense of conditions (12(e)–(g)), it is always possible [11] to find a function φ^* such that

$$\varphi(t) = \inf_w \{\varphi^*(t, w)\} \tag{13}$$

and such that φ^* is quadratic in t when w is fixed.

It can then be shown that the MAP criterion can be written as the minimum of a *dual* energy

$$J(f) = \inf_{b_x, b_y} J^*(f, b_x, b_y) \tag{14}$$

where J^* is given by

$$J^*(f, b_x, b_y) = \|p - Rf\|^2 + \lambda^2 \sum_k \varphi^*[(D_x f)_k, (b_x)_k] + \lambda^2 \sum_k \varphi^*[(D_y f)_k, (b_y)_k]. \tag{15}$$

Note that the dual energy is quadratic in f when b is fixed. This is why this kind of regularization is called "half-quadratic" regularization. Consequently, when the auxiliary variable is fixed, the first-order necessary conditions are linear in f.

B. Transforming the MAP Energy

The following theorem, proved in Appendix B, gives an implementation of the transformation in (13).

Theorem 1: Let φ be a potential function that satisfies conditions (12(a)) to (12(g)).

1) Then there exists a strictly convex and decreasing function $\psi: (0, M] \to [0, \beta)$, where

$$\beta = \lim_{t \to +\infty} \left(\varphi(t) - t^2 \frac{\varphi'(t)}{2t} \right)$$

such that

$$\varphi(t) = \inf_{0 < w \leq M} (wt^2 + \psi(w))$$

2) For every fixed $t \geq 0$, the value w_t for which the minimum is reached, i.e., such that

$$\inf_{0 < w \leq M} (wt^2 + \psi(w)) = \left(w_t t^2 + \psi(w_t)\right)$$

is unique and given by

$$w_t = \frac{\varphi'(t)}{2t}.$$

We give in Table II the analytical expression of ψ for three edge-preserving potentials.

Applying Theorem 1 to the MAP energy, with $t = (D_x f)_k$ and $w = (b_x)_k$ (respectively, with $t = (D_y f)_k$ and $w = (b_y)_k$), we obtain the following *augmented* energy:

$$J^*(f, b_x, b_y) = \|p - Rf\|^2 + \lambda^2 \sum_k \left\{(b_x)_k (D_x f)_k^2 + \psi[(b_x)_k]\right\} + \lambda^2 \sum_k \left\{(b_y)_k (D_y f)_k^2 + \psi[(b_y)_k]\right\}. \tag{16}$$

It is easy to verify that when the auxiliary variable is fixed, the augmented energy becomes quadratic in f, as stated in Section

III–A. But the augmented energy has another interesting property: Since ψ is convex in b (from Theorem 1), J^* is convex in b when f is fixed. Moreover, also from Theorem 1, the minimum value of b is unique and is given by

$$(b_x)_k = \frac{\varphi'(D_x f)_k}{2(D_x f)_k} \quad \text{and} \quad (b_y)_k = \frac{\varphi'(D_y f)_k}{2(D_y f)_k} \quad \forall k. \tag{17}$$

These properties are the basis for the reconstruction algorithm proposed in Section IV. Before describing the algorithm, let us make some additional remarks about Theorem 1.

First, since for any fixed t the minimizer of $\psi(w)$ is $w_t = \varphi'(t)/2t$, the central role of the weighting function is confirmed. According to the properties of this function, $(b_x)_k$ is close to zero for large gradients, and to M for small gradients (same remark for $(b_y)_k$). In other words, the value of b depends on the presence of an edge: b plays the role of a discontinuity marker, similar to the continuous "line process" defined by Geman and Reynolds in [17].

Second, it might be noticed that Theorem 1 is quite similar to the first theorem in [17]. However, there are two important differences. First, the Geman and Reynolds theorem only allows the introduction of the line variable when the potential function has a finite asymptotic behavior. Theorem 1 is also valid for certain potential functions that do not have an asymptote, and even for certain convex functions. Second, the expression for the minimum b is not explicitly given in [17]. Therefore, we think that Theorem 1 is a significant contribution to the theory of half-quadratic regularization.

In fact, there is a third more subtle difference between these theorems. We have supposed that the weighting function must be strictly decreasing (condition (12(e))). This implies that $\varphi(\sqrt{t})$ is supposed to be *strictly* concave, while in [17] it needs only be concave (even though the proof is given for the strictly concave case). As a consequence, Theorem 1 does not apply to some potential functions like the truncated quadratic or the Huber function [31]. This restriction, however, allows a one-to-one correspondence between the gradient and the auxiliary variable, which is important for the convergence proof of the following reconstruction algorithm.

IV. Deterministic Reconstruction Algorithm

A. Description of the Algorithm

As we have seen in the previous section, minimizing the MAP energy $J(f)$ in f is equivalent to minimizing $J^*(f,b)$ in (f,b). In order to exploit the properties of half-quadratic regularization, we propose to use a strategy based on alternate minimizations over f and b as follows:

$$\begin{aligned} &f^0 \equiv 0 \\ &\text{Repeat} \\ &\quad \left| \begin{array}{l} b^{n+1} = \arg\min_b [J^*(f^n, b)] \\ f^{n+1} = \arg\min_f [J^*(f, b^{n+1})] \end{array} \right. \\ &\text{Until convergence.} \end{aligned} \tag{18}$$

TABLE II
Three Edge-Preserving Potential Functions and the Corresponding Functions

Potential function	Expression of $\varphi(t)$	Expression of $\psi(w)$
φ_{GM}	$\frac{t^2}{1+t^2}$	$w - 2\sqrt{w} + 1$
φ_{HL}	$\log(1+t^2)$	$w - \log(w) - 1$
φ_{HS}	$2\sqrt{1+t^2} - 2$	$w + \frac{1}{w} - 2$

Since at step $n+1$, f^n is fixed, b^{n+1} is simply computed using the following expressions:

$$\left(b_x^{n+1}\right)_k = \frac{\varphi'\left[\left(D_x f^n\right)_k\right]}{2\left(D_x f^n\right)_k} \quad \text{and} \quad \left(b_y^{n+1}\right)_k \frac{\varphi'\left[\left(D_y f^n\right)_k\right]}{2\left(D_y f^n\right)_k} \ \forall k. \tag{19}$$

The minimization over f is also very simple since $J^*(f, b^{n+1})$ is quadratic. The new image estimate, f^{n+1}, is solution of the normal equations

$$\left(R^t R - \lambda^2 \Delta_A^{n+1}\right) f^{n+1} = R^t p \tag{20}$$

where $\Delta_A^{n+1} = -D_x^t B_x^{n+1} D_x - D_y^t B_y^{n+1} D_y, B_x^{n+1} = \operatorname{diag}[(b_x^{n+1})_k]$ and $B_y^{n+1} = \operatorname{diag}[(b_y^{n+1})_k]$. These equations can be solved by many iterative algorithms (cf. [30]), starting each step from the previous step's image estimate.

This algorithm, called ARTUR [8]–[10], solves the original nonquadratic minimization problem using a sequence of quadratic minimizations, which are easy to solve. In contrast to many reconstruction algorithms, this strategy can be applied for any potential function that satisfies the conditions for edge preservation (12).

Lastly, we would like to point out the importance of b_x and b_y in the preservation of edges. Since their value at each pixel site depends on the presence of an edge (through the weighting function), b_x and b_y play the role of discontinuity maps. At every step of the algorithm, new discontinuity maps are computed (from the last image estimate) and then taken into account for the computation of the new estimate. It can be observed (see Section V-A) that the discontinuity maps are very rough at first and become sharper as the algorithm is proceeding. Therefore, we can say that ARTUR is a "progressive discontinuity introduction" (PDI) algorithm.

B. Theoretical Study

The goal of this section is to give a justification to our algorithmic strategy. We have seen how the MAP energy can be transformed into a half-quadratic energy by setting

$$\min_f J(f) = \min_f \min_b J^*(f,b). \tag{21}$$

Since it is possible to reverse the order of the minimizations with respect to f and b, we have

$$\min_f J(f) = \min_f \min_b J^*(f,b) = \min_b \min_f J^*(f,b). \tag{22}$$

Now, let us define

$$T(b) = \min_f J^*(f, b). \tag{23}$$

Hence, we have

$$\min_f J(f) = \min_b T(b). \tag{24}$$

Since $J^*(f, b)$ is quadratic with respect to f when b is fixed, (23) has a unique solution, which we will denote f_b.

Let us also denote the minimizer of $T(b)$ by $\hat{b}$. It is possible to show (see Appendix C) the following results.

Theorem 2:

$$\text{if } \hat{b} = \arg\min_b T(b) \text{ then } f_{\hat{b}} = \arg\min_f J(f).$$

This result means that all minimizers of $T(b)$ yield image estimates that minimize the MAP energy.

It can be shown that these minimizers satisfy the following fixed-point equations

$$(\hat{b}_x)_k = \frac{\varphi'\left[\left(D_x f_{\hat{b}}\right)_k\right]}{2\left(D_x f_{\hat{b}}\right)_k} \quad \text{and} \quad (\hat{b}_y)_k = \frac{\varphi'\left[\left(D_y f_{\hat{b}}\right)_k\right]}{2\left(D_y f_{\hat{b}}\right)_k} \quad \forall k. \tag{25}$$

If the MAP energy is convex, then its minimum is unique and it is possible to show (see Section IV-C) that ARTUR converges to this solution.

C. Convergence of the Algorithm

Suppose that the potential function satisfies conditions (12). Then it is possible to demonstrate (see proof in Appendix D) the following convergence results.

Theorem 3: Let φ be a potential function that satisfies all conditions (12). Then

- the sequence $J_n = J^*(f^n, b^{n+1}) = J(f^n)$ is convergent;
- we have

  $$\left(b_x^{n+1} - b_x^n\right) \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{and} \quad \left(b_y^{n+1} - b_y^n\right) \underset{n\to+\infty}{\longrightarrow} 0$$
  $$\left(D_x f^n - D_x f^{n-1}\right) \underset{n\to+\infty}{\longrightarrow} 0$$

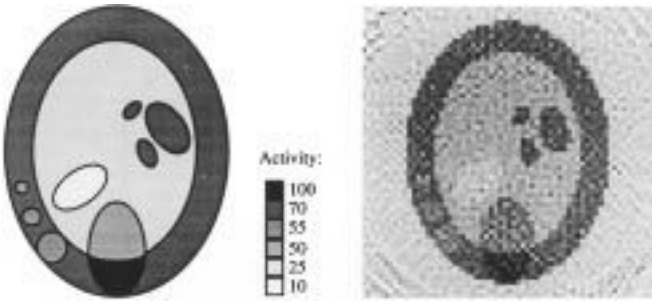
  and

  $$\left(D_y f^n - D_y f^{n-1}\right) \underset{n\to+\infty}{\longrightarrow} 0;$$

- if φ is convex on $[0, +\infty)$, then the sequences $Rf^n, D_x f^n, D_y f^n$ and, b^n are convergent. Moreover, if R is full rank, then the sequence f^n, is convergent.

When φ is strictly convex, the MAP energy can be strictly convex too. In fact, to ensure the strict convexity of the MAP energy, it is necessary that the null-spaces of R, D_x, and D_y do not intersect each other, which is generally the case. In this case, the minimum of the MAP energy is unique. The minimum in b of $T(b)$ is thus unique too, and there exists only one point where the derivative of $T(b)$ vanishes. When R is full rank, the sequence f^n converges. The computed estimate is then the unique solution of (2).

When φ is nonconvex, the first two results remain valid, in particular the convergence of the sequence J_n. However, the algorithm probably computes a local minimum of the MAP energy (see experimental results), which is not theoretically characterized yet.



(a) (b)

Fig. 4. (a) Synthetic phantom Nice and (b) unregularized solution after 50 iterations.

D. Remarks

First, as we wrote in Section II-A, we use in practice all two-pixel cliques of a second-order model. In particular, our complete model involves an additional term of the form $\lambda^2 J_d(f)$, with

$$J_d(f) = \sum_k \varphi[(D_{d1} f)_k] + \sum_k \varphi[(D_{d2} f)_k] \tag{26}$$

which is a function of *diagonal gradients*

$$(D_{d1} f)_{i,j} = (f_{i+1,j+1} - f_{i,j})/\delta\sqrt{2}$$

and

$$(D_{d2} f)_{i,j} = (f_{i+1,j-1} - f_{i,j})/\delta\sqrt{2}. \tag{27}$$

All the above theory remains valid for diagonal terms. The only modification to the algorithm is that the values of diagonal line variables are weighted by $1/2$

$$\left(b_{d1}^{n+1}\right)_k = \frac{1}{2}\frac{\varphi'\left[\left(D_{d1} f^n\right)_k\right]}{\left(2D_{d1} f^n\right)_k}$$

and

$$\left(b_{d2}^{n+1}\right)_k = \frac{1}{2}\frac{\varphi'\left[\left(D_{d2} f^n\right)_k\right]}{\left(2D_{d2} f^n\right)_k} \quad \forall k \tag{28}$$

Second, we would like to make a remark about positivity. In certain applications, as tomography, the solution is known to be positive. This is an important *a priori* information. In fact, a nonnegativity condition can be imposed on the solution by adding to the MAP energy a penalty term of the form $\gamma^2 J_+(f)$, with

$$J_+(f) = \sum_k \varphi_+(f_k), \ \varphi_+(t) = w_+ t^2$$

and

$$w_+ = \begin{cases} 1 & \text{if } t < 0 \\ 0 & \text{if } t \geq 0 \end{cases}. \tag{29}$$

This term involves a variable which marks negative values and which can be used in the same way as the discontinuity maps: At each step, a new negativity map is first computed, and then taken into account in the computation of the new estimate. Therefore, the introduction of a nonnegativity penalty only provokes slight modifications to the algorithm and gives good results.

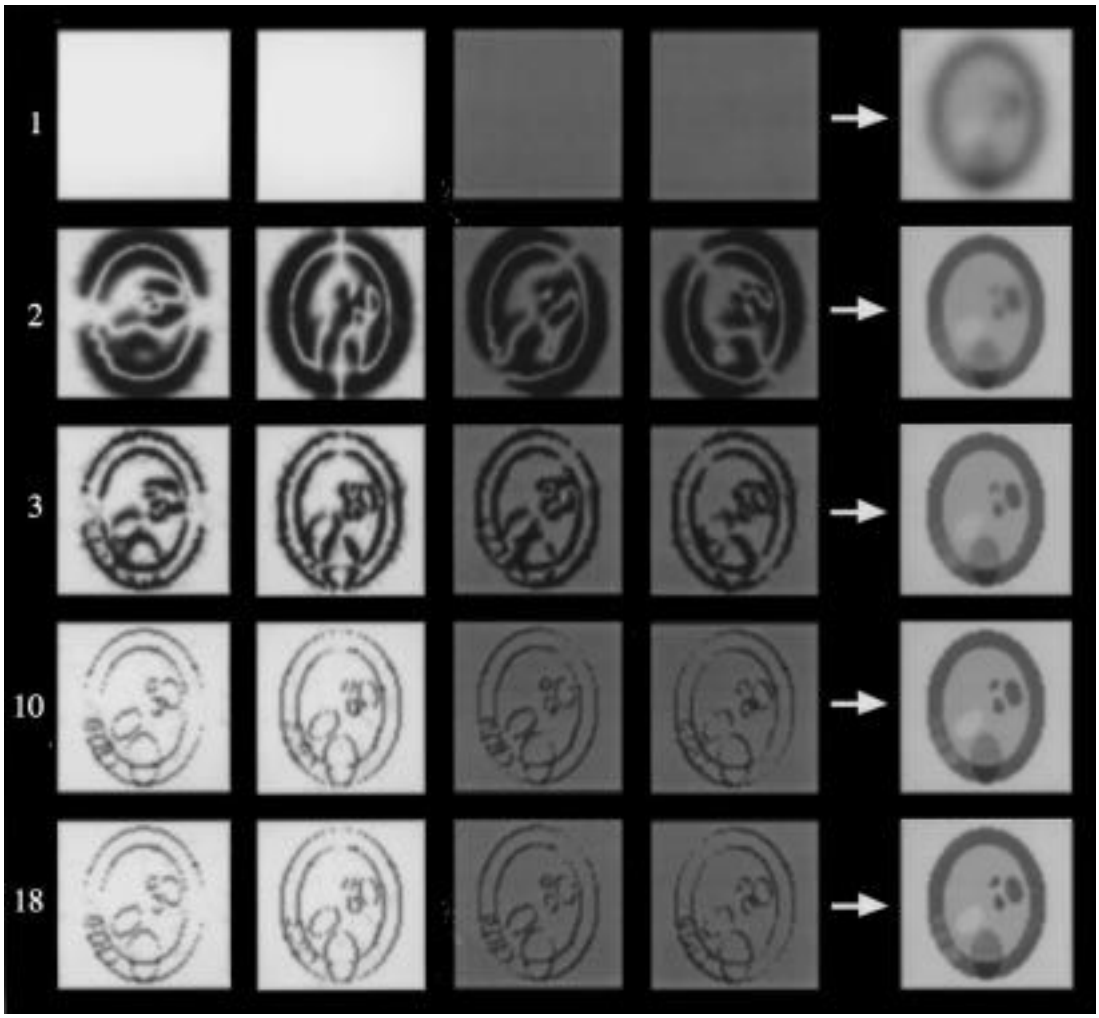


Fig. 5. Values of the auxillliary variables (from left to right: b_y, b_x, b_{d1}, and b_{d2}) and the image estimate after 1, 2, 3, 10, and 18 steps of the algorithm.

V. Experimental Results

In this section, we illustrate the behavior of ARTUR in the case of 2-D single photon emission tomography (SPECT) with both synthetic and real data.

A. Running of the Algorithm

We first present an example of reconstruction with synthetic data. The synthetic phantom named Nice is presented in Fig. 4(a). This phantom is a 64 $\times$ 64-pixel image derived from Shepp and Logan's head phantom, and models a cross-section of the brain with ellipses. Synthetic SPECT data are computed by projecting the 2-D phantom along 64 angles, on a 64-element detector. Projections in a certain direction are computed by summing the photons emitted by the object along this direction. Projections are then artificialy corrupted by Poisson noise. The resulting total number of counts (i.e., number of detected photons) in the projection vector is about 6 000 000. The signal-to-noise ratio (SNR) on projection vector is 26.7 dB (variances ratio). Undoubtedly, 6 000 000 counts is an order of magnitude too high for SPECT, but our aim in this section is just to illustrate the running of the algorithm. Therefore, we take a very favorable situation. An example of reconstruction with real data (250 000 counts) is given in Section V-B.

Fig. 4(b) shows an unregularized solution obtained by iteratively minimizing $J_1(f)$, as defined in (4), using a Gauss–Seidel algorithm. The algorithm has been stopped after 50 iterations. The SNR between the reconstructed image and the original one is 8.18 dB (variances ratio). This illustrates the noise amplification effect due to the ill posedness of the problem.

Fig. 5 shows a regularized reconstruction with ARTUR. For this reconstruction, we use φ_{GM}, which is nonconvex. The parameters of the model are experimentally fixed to $\lambda^2 = 525$ and $\delta = 7$. We let ARTUR proceed until the relative norm

$$\|f^{n+1} - f^n\|^2 / \|f^n\|^2$$

becomes smaller than $1e$-6 (n being the step number). In this case, the normal equations at each step are solved using a Gauss–Seidel algorithm, which stops when the relative norm

$$\|f^{n+1,m+1} - f^{n+1,m}\|^2 / \|f^{n+1,m}\|^2$$

becomes smaller than $1e$-6 (m being the iteration number). The algorithm stopped after 18 steps, for a total number of 104 iterations (processing time: 10 second per iteration).

Note that other algorithms than Gauss–Seidel can be used to solve the normal equations at each step. For example, the same reconstruction with a conjugate gradient (CG) algorithm needs about 80 iterations, and the reconstruction time is 1 s per iteration for a 64×64 image on a DEC-5000/240 workstation.

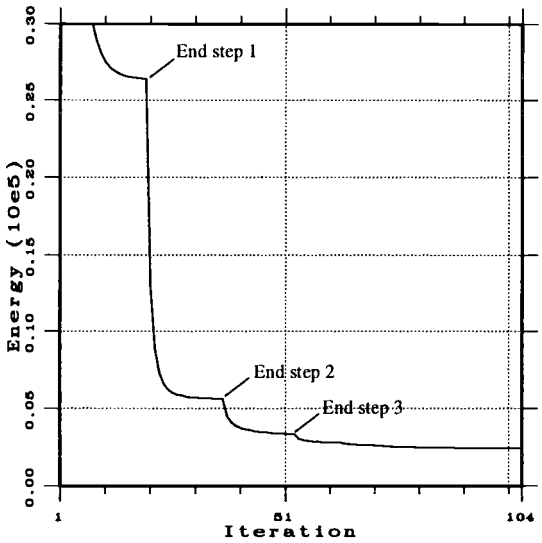
The evolution of both auxiliary variables and image estimates are shown in Fig. 5 after $n = 1, 2, 3, 10$ and 18 steps. For a better visualization, auxiliary variables are plotted in grey levels, while image estimates are in inverse grey levels (the colormap is given in Fig. 4). It can be noticed that diagonal line variables appear in darker grey than horizontal and vertical variables. This is because they are weighted by $1/2$, as remarked in Section IV-3.

It can also be observed in Fig. 5 that all auxiliary variables are homogeneous at step 1. This is normal since they are computed from the original guess, which is a uniformly null image. Then b^1 is uniformly equal to 1 and the first image estimate then corresponds to the solution that would be estimated with a Tikhonov regularization (i.e., using a quadratic potential), with regularization parameter equal to 525, which is a high value in this case. This explains the very smooth aspect of the first estimate. At step 2, new values of the auxiliary variables are computed using the first image estimate, and utilized to calculate the new image estimate, and so on.

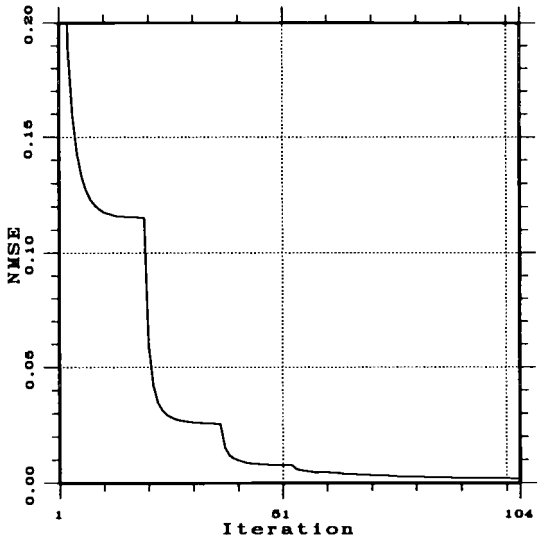
Note that other initial guesses could be used [10]. However, we prefer using a null image because this allows an important noise elimination during the first step. We think that this explains the good results we obtain in the nonconvex case.

The role of discontinuity maps played by the auxiliary variables clearly appears in Fig. 5. Since they are computed from smooth images, the first discontinuity maps are very rough. As the algorithm proceeds, they become more precise. In fact, the joint estimation of the image and its discontinuity maps progresses as follows. At each step, discontinuities are introduced into the new image estimate. The new discontinuity maps that are computed from this image are then sharper. More discontinuities are then introduced into the image estimate, and so on.

In Fig. 6, we show plots of the MAP energy, J, and of the NMSE between the original and estimated images versus the iteration number. They both illustrate the convergence properties of ARTUR. The left-hand curve confirms the first result of Theorem 3, namely the fact that the sequence of the MAP energies at the beginning of each step is convergent. It can also be observed that most of the "work" of the algorithm is concentrated into the first steps, as it can also be seen in Fig. 5.



(a)



(b)

Fig. 6. Evolution of the MAP energy (left-hand curve) and of the NMSE between the estimate and the original image across the iterations.

B. Examples of Reconstruction with Different Potential Functions

In Figs. 7 to 10, we give examples of reconstruction with different potentials, all satisfying (12) except for the quadratic function $\varphi_Q(t) = t^2$. Our aim is not to provide a rigorous comparison between these potential functions—in fact this would require more experiments and the definition of quality criteria—but to give some indications about their behavior.

Fig. 7 shows reconstructions of the synthetic phantom Nice. Profiles of these reconstructed images are shown in Fig. 8. The first idea these figures illustrate is the superiority of edge-preserving models over quadratic regularization for reconstructing piecewise constant images. The second idea is that nonconvex potentials seem to provide better reconstructions than convex potentials. More precisely, nonconvex potentials seem to yield sharper edges. Of course, the convergence of ARTUR to the solution of the MAP is not theoretically proved. However, we find that the computed (local) minimum is visually very satisfying.

The same observations about the quality of reconstructed images can be made about Fig. 10, that shows reconstructions of a real Jaszczak phantom, shown in Fig. 9. The data were acquired in clinical conditions, the total number of counts being about 250 000 for the considered cross-section. Note that no physical correction was made in this case. This is why the center of the reconstructed object is not homogeneous. Also, it can be noticed that ARTUR yields better results than

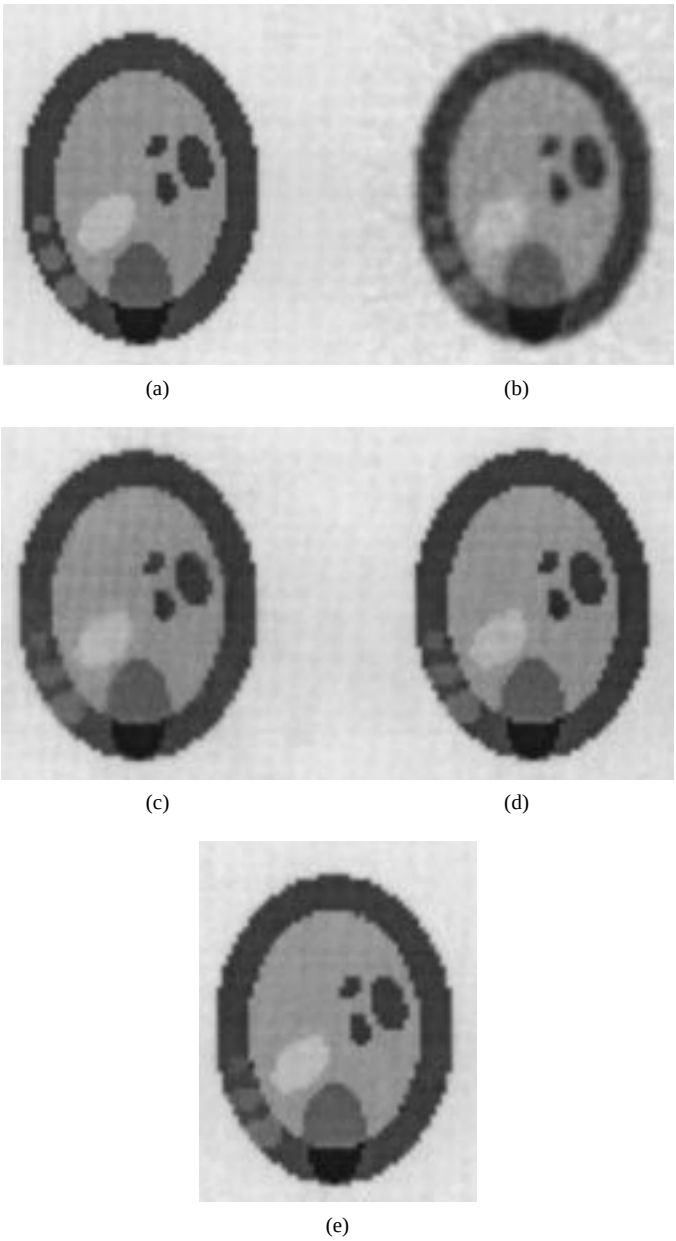


Fig. 7. (a) Synthetic object and (b) reconstructed images with different potential functions as follows: (b) $\theta_Q(t) = t^2$; (c) θ_{GR}; (d) θ_{HL}; (e) θ_{GM}.

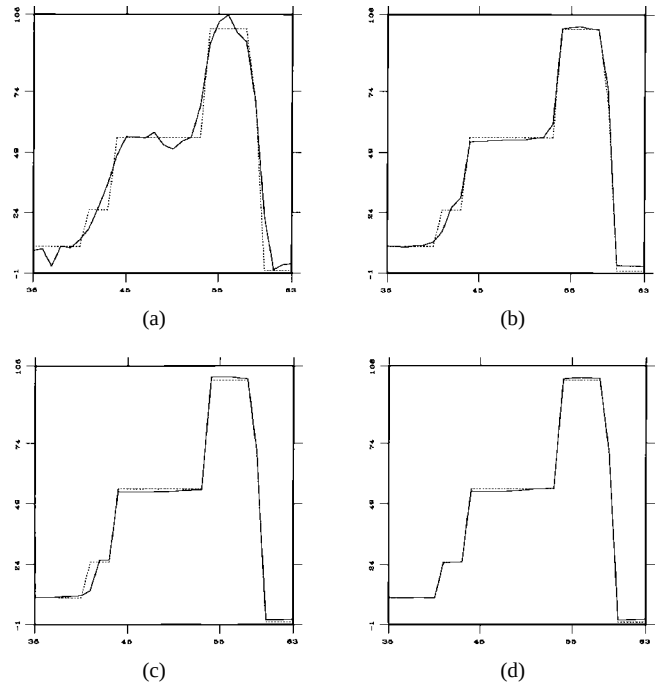


Fig. 8. Profiles of the synthetic object (dotted line) and reconstructed images (solid line) with (a) $\theta_Q(t) = t^2$; (b) θ_{GR}; (c) θ_{HL}; and (d) θ_{GM}.

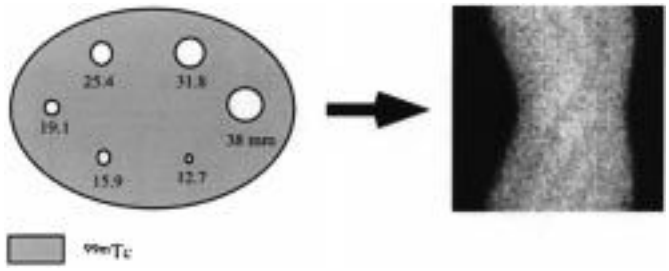


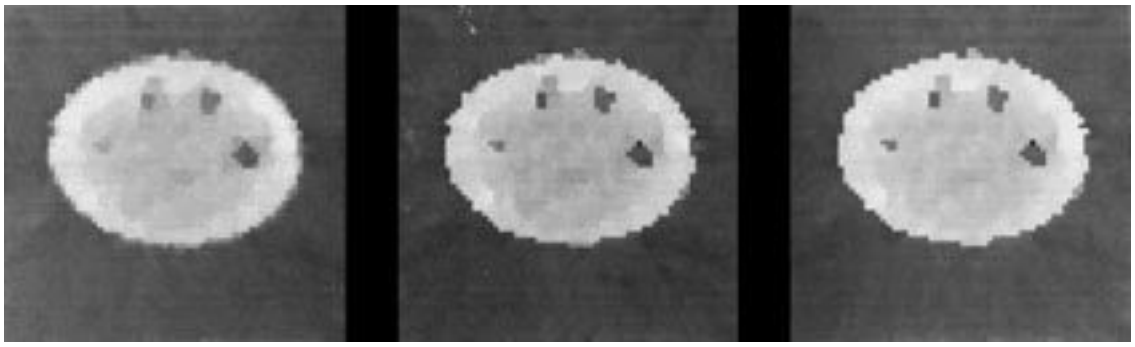
Fig. 9. Jaszczak phantom, left, and its projections, right.

the usual convolution back-projection (CBP) algorithm. The results obtained by ARTUR with φ_{GM} are comparable to those obtained by Green's MAP expectation-maximization "one step late" (MAP-EM-OSL) [20] algorithm. With the same potential (see Figs. 10(d) and (c)), ARTUR yields better results than the MAP-EM-OSL algorithm, for a lower total reconstruction time [25].

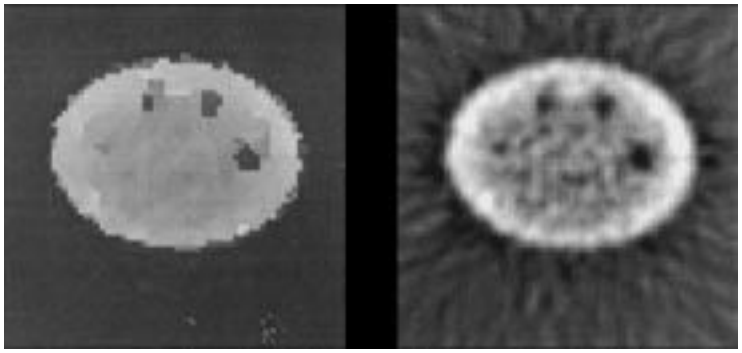
VI. Conclusion

In this paper, we have considered the problem of edge-preserving regularization in computed imaging. Our first aim was to give a unified answer to the question, "What properties must a potential function (or its derivative) satisfy to ensure the preservation of edges?" We have proposed a heuristical study of the first-order necessary conditions which led us to propose three *conditions for edge preservation*. These conditions are not imposed on the potential function, φ, but on the *weighting function*, $\varphi'(t)/2t$. They are satisfied by many of the functions that can be found in the literature. A consequence is that edge preservation can be performed using functions that do not have a finite asymptotic behavior and even by convex potentials. This is important since strict convexity generally ensures uniqueness of the solution. We also propose a convex differentiable and practical edge-preserving potential, φ_{HS}.

The second contribution of this paper concerns half-quadratic regularization. We have shown that when the potential function satisfies the conditions for edge preservation, it is possible to introduce an auxiliary variable, b, whose role is twofold: to mark discontinuities (b corresponds in fact to the continuous "line process" defined in [17]), and to allow the linearization of the problem. When b is fixed, the energy becomes quadratic with respect to f. On the other hand, when f is fixed, the energy is convex with respect to the auxiliary variable. Moreover, in this case Theorem 1 gives an explicit expression for the b that minimizes the energy.



(a) (b) (c)



(d) (e)

Fig. 10. Reconstructions of the Jaszczak phantom. Top row: ARTUR with (a) θ_{GR}; (b) θ_{HL}; (c) $\backslash\theta_{GM}$. Bottom row: MAP-EM-OSL with (d) θ_{GM}; (e) CBP.

We have exploited the properties of half-quadratic regularization to derive a new adaptive deterministic reconstruction algorithm based on alternate minimizations on f and b, called ARTUR, which is a progressive discontinuity introduction (PDI) algorithm. The PDI principle is very general in the sense that it can be applied to any edge-preserving function in the sense of (12), in many reconstruction problems (e.g., Computed tomography, stereovision, restoration, motion estimation). It is very simple to implement since it is based on quadratic minimizations. Finally, it is efficient: the convergence to the unique minimum of the MAP is proved in the convex case, and the results are very satisfactory in the nonconvex case. Note that the PDI principle has already been used to develop two other reconstruction algorithms. The first one, called LEGEND [1], [10], [11], corresponds to the case where the principle and auxiliary variables interact with each other in an additive way (in this paper we have considered a multiplicative interaction). The second one is called MOISE [10]. In this case, we consider a Poisson noise model instead of the Gaussian model in (1).

Our future work on this subject will follow three main directions. First, we have seen that even if the PDI principle can be applied to a large variety of functions, the quality of reconstructed images depends on the potential. It would be interesting to make theoretical and practical comparisons between different potentials. This work is initiated in [5]. A second perspective of work would be the quantitative comparison between ARTUR and other existing reconstruction algorithms in both terms of quality and computational efficiency. Finally, we are adapting ARTUR to the three-dimensional (3-D) case, in order to make 3-D reconstruction accessible to most of the usual workstations.

APPENDIX A
FIRST-ORDER NECESSARY CONDITIONS

A minimum of J necessarily satisfies $\partial J/\partial f = 0$. The calculus of $\partial J_1/\partial f$ is straightforward. Now, let us calculate the local expression for $\partial J_2/\partial J$ at site (i,j). To simplify, we suppose that site (i,j) is not located at the border of the image. The case of boundaries needs a special treatment that will not be discussed here. We have then

$$\left(\frac{\partial J_2}{\partial f}\right)_{i,j} = -\varphi'(f_{i,j+1} - f_{i,j}) + \varphi'(f_{i,j} - f_{i,j-1}) - \varphi'(f_{i+1,j} - f_{i,j}) + \varphi'(f_{i,j} - f_{i-1,j}).$$

Supposing that φ satisfies (12g), it is possible to rewrite this expression as

$$\left(\frac{\partial J_2}{\partial f}\right)_{i,j} = 2\Bigg[-(f_{i,j+1} - f_{i,j})\frac{\varphi'(f_{i,j+1} - f_{i,j})}{2(f_{i,j+1} - f_{i,j})} + (f_{i,j} - f_{i,j-1})\frac{\varphi'(f_{i,j} - f_{i,j-1})}{2(f_{i,j} - f_{i,j-1})} - (f_{i+1,j} - f_{i,j})\frac{\varphi'(f_{i+1,j} - f_{i,j})}{2(f_{i+1,j} - f_{i,j})} + (f_{i,j} - f_{i-1,j})\frac{\varphi'(f_{i,j} - f_{i-1,j})}{2(f_{i,j} - f_{i-1,j})}\Bigg]$$

or

$$\left(\frac{\partial J_2}{\partial f}\right)_{i,j} = -2\Big\{\lambda^{\mathcal{E}} f_{i,j+1} + \lambda^{\mathcal{W}} f_{i,j-1} + \lambda^{\mathcal{S}} f_{i+1,j} + \lambda^{\mathcal{N}} f_{i-1,j} - \sum f_{i,j}\Big\}$$

with the definitions given in Fig. 2. In other words, the derivative of J_2 at site (i,j) is obtained by filtering the image f by the (position dependent) filter shown in Fig. 2, which is a 3×3 weighted Laplacian. If all weights are equal to 1, this filter is the usual 3×3 Laplacian filter. This can be written as a matrix-vector multiplication, and then we obtain the following expression for the necessary conditions

$$\frac{\partial J}{\partial f} = 0 \Leftrightarrow \left(R^t R - \lambda^2 \Delta_{\text{pond}}\right) f = R^t p$$

where Δ_{pond} is the matrix corresponding to the weighted discrete approximation of the Laplacian. When all weights are equal to 1, we have $\Delta_{\text{pond}} = \Delta$, where Δ is the matrix corresponding to the usual Laplacian operator.

Appendix B
Proof of Theorem 1

The proof is similar to the one of Geman and Reynolds [17]. Let us define $\theta(t) = \varphi(\sqrt{t})$. Since φ satisfies (12(e)), it is straightforward that $\theta(t)$ is strictly concave. Therefore, we have

i) $\theta'(t) = \frac{\varphi'(\sqrt{t})}{2\sqrt{t}}$, is a strictly decreasing function $(\theta' : [0, +\infty) \to (0, M])$

ii) $\forall u, v$, in $[0, +\infty), \theta(u) \leq \theta(v) + (u - v)\theta'(v)$ with equality only if $u = v$.

From i), it is clear that $\theta'(t)$ is one-to-one and admits an inverse: $(\theta')^{-1}(w) : (0, M] \to [0, +\infty)$. From ii), we deduce that

$$\theta(u) = \inf_v \{\theta'(v)u + \theta(v) - v\theta'(v)\}. \tag{B1}$$

1) Let us define $w = \theta'(v)$ or equivalently $v = (\theta')^{-1}(w)$ and $\psi(w) = \theta((\theta')^{-1}(w)) - w(\theta')^{-1}(w)$. Then, (B1) becomes

$$\theta(u) = \inf_w \{wu + \psi(w)\}.$$

Letting $u = t^2$, we obtain

$$\theta(t^2) = \varphi(t) = \inf_w \{wt^2 + \psi(w)\}.$$

A simple calculation shows that $\psi'(w) = -(\theta')^{-1}(w)$. Since $(\theta')^{-1}$ takes its values in $[0, +\infty)$, this shows that ψ is strictly decreasing on $(0, M]$. Hence, its limits at 0^+ and $+\infty$ exist. From the definition of ψ, we have on one hand

$$\alpha = \lim_{w \to M} \psi(w) = \lim_{v \to 0^+} (\theta(v) - v\theta'(v)).$$

From (12(a)), we have $\theta(0) = 0$. From (12(f)), we also have $\theta'(0) = M < +\infty$. Thus, clearly, $\alpha = 0$. On the other hand

$$\beta = \lim_{w \to 0^+} \psi(w) = \lim_{v \to +\infty} (\theta(v) - v\theta'(v)).$$

Letting $v = t^2$ we obtain the definition of β given in Theorem 1. Note that β is not necessarily finite.

Finally, the expression of $\psi''(w)$ is

$$\psi''(w) = \lim_{h \to 0} \frac{\psi'(w+h) - \psi'(w)}{h} = \lim_{h \to 0} \frac{-\theta'^{-1}(w+h) + \theta'^{-1}(w)}{(w+h) - (w)}$$

which can be written as

$$\psi''(w) = -\lim_{h \to 0} \frac{\theta'^{-1}(w) - \theta'^{-1}(w+h)}{\theta'[\theta'^{-1}(w)] - \theta'[\theta'^{-1}(w+h)]}.$$

Letting $v = \theta'^{-1}(w)$ and $e = \theta'^{-1}(w+h)$, we obtain

$$\psi''(w) = -\lim_{e \to v} \frac{v - e}{\theta'(v) - \theta'(e)}.$$

Hence, we have $\psi''(w) = \frac{-1}{\theta''(v)}$, which is strictly positive. Thus, ψ is strictly convex.

2) From (B1), it is obvious that if u is fixed, then the minimum is reached for $v = u$, i.e., for $w = \theta'(v)$. With $v = t^2$ we have then $w = \frac{\varphi'(t)}{2t}$.

Appendix C
Proof of Theorem 2

By construction

$$J(f) = \min_b J^*(f, b) \quad \forall f$$

so

$$J(f) \leq J^*(f, b) \quad \forall f \ \forall b.$$

In particular

$$J(f_{\hat{b}}) \leq J^*\left(f_{\hat{b}}, \hat{b}\right) \tag{C1}$$

Now, since by definition of $\hat{b}$: $T(\hat{b}) \leq T(b)\ \forall b$, from the definition of $T(b)$ it is clear that

$$J^*\left(f_{\hat{b}}, \hat{b}\right) \leq \min_f J^*(f, b) \quad \forall b.$$

So in particular

$$J^*\left(f_{\hat{b}}, \hat{b}\right) \leq \min_b \min_f J^*(f, b) = \min_f \min_b J^*(f, b) = \min_f J(f). \tag{C2}$$

Combining (C1) and (C2) yields

$$J(f_{\hat{b}}) \leq \min_f J(f).$$

So $f_{\hat{b}}$ minimizes the MAP energy, $J(f)$.

APPENDIX D
PROOF OF THEOREM 3

First, let us consider the sequence $J_n = J^*(f^n, b^{n+1})$. From (14), we have $J(f^n) = \min_b J^*(f^n, b)$ From the definition of ARTUR (principle of alternate minimizations), b^{n+1} is the minimizer of $J^*(f^n, b)$. Hence we have $J_n = J^*(f^n, b^{n+1}) = J(f^n)$.

Now, let us show that it is convergent. Since b^{n+1} minimizes $J^*(f^n, b)$, it satisfies

$$J^*(f^n, b^{n+1}) \leq J^*(f^n, b) \quad \forall b, \ \forall n. \tag{D1}$$

On the other hand f^n minimizes $J^*(f, b^n)$

$$J^*(f^n, b^n) \leq J^*(f, b^n) \quad \forall f, \ \forall n. \tag{D2}$$

We have

$$\begin{aligned} J_{n-1} - J_n = &[J^*(f^n, b^n) - J^*(f^n, b^{n+1})] \\ &+ [J^*(f^{n-1}, b^n) - J^*(f^n, b^n)]. \end{aligned} \tag{D3}$$

Applying (D1) for $b = b^n$ and (D2) for $f = f^{n-1}$, we obtain

$$J^*(f^n, b^n) - J^*(f^n, b^{n+1}) \geq 0 \tag{D4}$$

$$J^*(f^{n-1}, b^n) - J^*(f^n, b^n) \geq 0. \tag{D5}$$

From (D4) and (D5), we deduce that J_n is decreasing

$$J_{n-1} - J_n \geq 0 \quad \forall n.$$

Moreover, it is clear that $J_n(f) \geq 0 \ \forall f$ because $\varphi(t) \geq 0 \ \forall t$. Since it is decreasing and bounded below, the sequence J_n is convergent.

Now, let us show that $[(b^{n+1})_k - (b^n)_k]_{n \to +\infty} \to 0 \ \forall k$. For a better readability, we consider hereafter the one-dimensional (1-D) case. We note b the line process and D the derivation operator.

Let us calculate

$$\begin{aligned} &J^*(f^n, b^n) - J^*(f^n, b^{n+1}) \\ &\quad = \lambda^2 \sum_k \left\{ (b^n)_k (Df^n)_k^2 + \psi\left[(b^n)_k\right] \right\} \\ &\qquad - \lambda^2 \sum_k \left\{ (b^{n+1})_k (Df^n)_k^2 + \psi\left[(b^{n+1})_k\right] \right\}. \end{aligned} \tag{D6}$$

Let us define $g_k^n(w) = w(Df^n)_k^2 + \psi(w)$. Using a Taylor development of $g_k^n(\cdot)$, we have

$$\begin{aligned} &J^*(f^n, b^n) - J^*(f^n, b^{n+1}) \\ &\quad = \lambda^2 \sum_k \left(\left[(b^n)_k - (b^{n+1})_k\right] \left(g_k^n\right)' \left[(b^{n+1})_k\right] \right) \\ &\qquad + \frac{\lambda^2}{2} \sum_k \left(\left[(b^n)_k - (b^{n+1})_k\right]^2 \left(g_k^n\right)'' \left(c_k^n\right) \right) \end{aligned} \tag{D7}$$

with $c_k^n \in (b_k^n, b_k^{n+1})$. Since b^{n+1} minimizes $J^*(f^n, b)$, we have

$$\left(g_k^n\right)' \left[(b^{n+1})_k\right] = (Df^n)_k^2 + \psi'\left[(b^{n+1})_k\right] = 0.$$

Consequently

$$\begin{aligned} &J^*(f^n, b^n) - J^*(f^n, b^{n+1}) \\ &\quad = \frac{\lambda^2}{2} \sum_k \left(\left[(b^n)_k - (b^{n+1})_k\right] \left(g_k^n\right)'' \left(c_k^n\right) \right). \end{aligned} \tag{D8}$$

We have $(g_k^n)''(w) = \psi''(w)$. We also know (see Appendix B) that $\psi''(w) = -1/\theta''(t^2)$. With $w = \theta'(t^2) = \varphi'(t)/2t$, a simple calculus shows that

$$-\theta''(t^2) = \frac{1}{2t^2} \left(\frac{\varphi'(t)}{2t} - \frac{\varphi''(t)}{2} \right).$$

Since this function is continuous, it is bounded on any bounded interval $[t_1, t_2]$, with $t_1 > 0, t_2 < +\infty$. We know from (12e–f) that $\varphi'(t)/2t$ is continuous and strictly decreasing and that its limit when $t \to +\infty$ is zero, so its derivative is negative and necessarily vanishes when $t \to +\infty$. Thus

$$\begin{aligned} \lim_{t \to +\infty} \left(\frac{\varphi'(t)}{2t} \right)' &= \lim_{t \to +\infty} \frac{1}{2t} \left(\frac{\varphi'(t)}{t} - \varphi''(t) \right) = 0 \\ &= \lim_{t \to +\infty} \frac{1}{2t^2} \left(\frac{\varphi'(t)}{t} - \frac{\varphi''(t)}{2} \right). \end{aligned}$$

Now, for $t = 0$, a Taylor development yields

$$\frac{1}{2t^2} \left(\frac{\varphi'(t)}{2t} - \frac{\varphi''(t)}{2} \right) = -\frac{1}{12} \varphi^{(4)}(0) + \varepsilon(t)$$

(remember that condition (12(h)) is $\varphi'''(0) = 0$, and note that conditions (12(b)–12(c)) imply $\varphi'(0) = 0$).

Thus, $-\theta''(t^2)$ can be bounded above for all values of t, i.e., $\psi''(w)$ can be bounded below. Then, $\exists\, c > 0$ such that

$$\left(g_n^k\right)''(w) \geq c \quad \forall w.$$

Then, finally

$$J^*(f^n, b^n) - J^*(f^n, b^{n+1}) \geq \frac{1}{2} \lambda^2 c \|b^n - b^{n+1}\|^2$$

and, accordingly to (D3) and (D5)

$$J_{n-1} - J_n \geq \frac{1}{2} \lambda^2 c \|b^n - b^{n+1}\|^2.$$

As the sequence J_n is convergent, we conclude that, $(b^{n+1} - b^n) \underset{n \to +\infty}{\to} 0$ is convergent too. Also, as $\varphi'(u)/2u$ is one-to-one, we have: $(Df^n - Df^{n-1}) \underset{n \to +\infty}{\to} 0$.

Finally, let us show that, in the convex case, the sequence f^n converges and thus that ARTUR finds the global minimizer of $J(f)$.

When b^n is fixed, $J^*(f, b^n)$ is convex in f and the minimum is attained for $f = f^n$. Thus

$$(R^t R + \lambda^2 D^t B^n D) f^n - R^t p = 0 \tag{D9}$$

with $B^n = \operatorname{diag}\{(b^n)_k, \ k = 1 \cdots N^2\}$ and $(b^n)_k = \frac{\varphi'[(Df^{n-1})_k]}{2(Df^{n-1})_k} \ \forall k$.

On the other hand, the minimum of $J, f_{\hat{b}}$, is unique and satisfies the first-order necessary conditions

$$(R^t R + \lambda^2 D^t \hat{B} D) f_{\hat{b}} - R^t p = 0 \tag{D10}$$

with $\hat{B} = \operatorname{diag}\{(\hat{b})_k, \ k = 1 \cdots N^2\}$ and $(\hat{b})_k = \frac{\varphi'[(Df_{\hat{b}})_k]}{2(Df_{\hat{b}})_k} \ \forall k$.

Subtracting (D10) to (D9) and making the scalar product with $(f^n - f_{\hat{b}})$, we obtain

$$\left\|R(f^n - f_{\hat{b}})\right\|^2 + \lambda^2\left\langle B^n Df^n - \hat{B}Df_{\hat{b}}, D(f^n - f_{\hat{b}})\right\rangle = 0.$$

Let us introduce $B^{n+1}Df^n$

$$\left\|R(f^n - f_{\hat{b}})\right\|^2 + \lambda^2\left\langle B^{n+1}Df^n - \hat{B}Df_{\hat{b}}, D(f^n - f_{\hat{b}})\right\rangle + \lambda^2\left\langle (B^n - B^{n+1})Df^n, D(f^n - f_{\hat{b}})\right\rangle = 0. \quad \text{(D11)}$$

Let us recall that

$$(b^{n+1})_k(Df^n)_k = \frac{1}{2}\varphi'\left[(Df^n)_k\right]$$

and

$$(\hat{b})_k(Df_{\hat{b}})_k = \frac{1}{2}\varphi'\left[(Df_{\hat{b}})_k\right].$$

Since φ is supposed to be convex (i.e., $[\varphi'(u) - \varphi'(v)] \cdot (u - v) \geq 0$) we have

$$\left\langle (B^{n+1})(Df^n) - (\hat{B})(Df_{\hat{b}}),\ (Df^n) - (Df_{\hat{b}})\right\rangle \geq 0.$$

Consequently, from (D11) we have

$$\left\|R(f^n - f_{\hat{b}})\right\|^2 - \lambda^2\left\langle (B^{n+1} - B^n)Df^n, D(f^n - Df_{\hat{b}})\right\rangle \leq 0$$

or

$$\left\|R(f^n - f_{\hat{b}})\right\|^2 \leq \lambda^2\|B^{n+1} - B^n\| \cdot \left[\|Df^n\| \cdot \left\|D(f^n - f_{\hat{b}})\right\|\right].$$

It is possible to show that the norm of the gradient is always bounded above by a constant, G. Then we have

$$\left\|R(f^n - f_{\hat{b}})\right\|^2 \leq 2\lambda^2 G^2\|B^{n+1} - B^n\|$$

and we conclude that $R(f^n - f_{\hat{b}}) \underset{n\to+\infty}{\to} 0$ because $(b^{n+1} - b^n) \underset{n\to+\infty}{\to} 0$. If R is full rank, it follows that $f^n \underset{n\to+\infty}{\to} f_{\hat{b}}$.

- Even if R is not full rank, we can show that $Df^n \underset{n\to+\infty}{\to} Df_{\hat{b}}$.. According to (D11), we have

$$\left\langle B^{n+1}Df^n - \hat{B}Df_{\hat{b}},\ D(f^n - f_{\hat{b}})\right\rangle \to 0.$$

This expression can be rewritten as

$$\sum_k \left\{\left[\varphi'(Df^n)_k\right] - \left[\varphi'(Df_{\hat{b}})_k\right]\right\}\left\{(Df^n)_k - (Df_{\hat{b}})_k\right\} \to 0. \quad \text{(D12)}$$

To finish the proof, we use a standard result in convex analysis. For a proof of this theorem, the reader is refered to [19].

Theorem: Let $J: \Re^n \to \Re$ be a strictly convex and C^1 function, then $\forall K > 0$, there exists a function $\rho_K: [0, 2K] \to \Re^+$, continuous, strictly increasing with $\rho_K(0) = 0$ and such that

$$\langle J'(v) - J'(u), v - u\rangle \geq \rho_K(\|v - u\|)$$
$$\forall u, v \text{ such that } \|u\| \leq K \text{ and } \|v\| \leq K.$$

Applying this theorem with $J(t) = \sum_k \varphi(t_k), v = Df^n, u = Df_{\hat{b}}$, and $K = \max(G, \|Df_{\hat{b}}\|)$, we know that there exists a function ρ_K such that

$$\sum_k \left\{\left[\varphi'(Df^n)_k\right] - \left[\varphi'(Df_{\hat{b}})_k\right]\right\} \times \left\{(Df^n)_k - (Df_{\hat{b}})_k\right\} \geq \rho_K\left(\left\|Df^n - Df_{\hat{b}}\right\|\right). \quad \text{(D13)}$$

It is clear that (D12) and (D13) imply

$$\lim_{n\to+\infty} \rho_K\left(\left\|Df^n - Df_{\hat{b}}\right\|\right) = 0.$$

But ρ_K is strictly increasing with $\rho_K(0) = 0$. Thus, we have necessarily $\lim_{+\infty}\|Df^n - Df_{\hat{b}}\| = 0$ and we conclude that $Df^n \underset{n\to+\infty}{\to} Df_{\hat{b}}$.

ACKNOWLEDGMENT

The authors thank J. Darcourt and O. Migneco from Nice Faculty of Medicine for their fruitful collaboration and for providing us with real data. They also thank two anonymous reviewers for their very careful reading of the manuscript and their sound advice.

REFERENCES

[1] G. Aubert, L. Blanc-Féraud, M. Barlaud, and P. Charbonnier, "A deterministic algorithm for edge-preserving computed imaging using Legendre transform," in *Proc. 12th Int. Conf. on Pattern Recognition*, Jerusalem, Israël, Oct. 1994, pp. 188–191.

[2] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat.. Soc.*, pp. 259–302, 1986.

[3] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[4] L. Blanc-Féraud, M. Barlaud, and T. Gaidon, "Motion estimation involving discontinuities in a multiresolution scheme," *Opt. Eng.*, vol. 32, pp. 1475–1482, July 1993.

[5] L. Blanc-Féraud, P. Charbonnier, G. Aubert, and M. Barlaud, "Nonlinear image processing: Modeling and fast algorithms for edge preservation," in *Proc. 2nd IEEE Int. Conf. on Image Processing*, Washington, DC Oct. 1995.

[6] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[7] B. Chalmond, "Image restoration using an estimated Markov model," *Signal Processing*, vol. 15, pp. 115–129, 1988.

[8] P. Charbonnier, L. Blanc-Féraud, and M. Barlaud, "ARTUR: An adaptive deterministic relaxation algorithm for edge-preserving tomographic reconstruction," Res. Rep. 93-76, I3S, Univ. Nice, Sophia Antipolis, France, 1993.

[9] ——— "An adaptive reconstruction method involving discontinuities," in *Proc. IEEE ICASSP 1993*, vol. V, Minneapolis, MN, pp. 491–494.

[10] P. Charbonnier, "Reconstruction d'image: Régularization avec prise en compte des discontinuités," Ph.D. dissertation, Univ. Nice, Sophia Antipolis, France, Sept. 1994.

[11] P. Charbonnier, G. Aubert, L. Blanc-Féraud, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. 1st IEEE ICIP*, Austin, TX, Nov. 1994.

[12] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, Dec. 1989.

[13] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, June 1994.

[14] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 401–412, May 1991.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[16] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Statistical Computation Section, Amer. Statistical Assoc.*, Washington, DC, 1985, pp. 12–18.

[17] S. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, Mar. 1992.

[18] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization and FFT's," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, July 1995.

[19] R. Glowinski, J. L. Lions, and R. Tremolières, *Analyze numérique des inéquations variationnelles, Tome 1: Théorie Générale, Méthodes mathématiques pour l'informatique*. Paris, France: Dunod, 1976.

[20] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. 9, pp. 84–93, Mar. 1990.
[21] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. MI-8, pp. 194–202, June 1990.
[22] B. L. P. Horn, *Robot Vision* Cambridge, MA: MIT Press, 1986.
[23] A. K. Jain, *The Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1989.
[24] F. C. Jeng and J. W. Woods, "Compound Gauss–Markov random fields for image estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 683–697, Mar. 1991.
[25] P. M. Koulibaly, P. Charbonnier, M. Barlaud, J. Darcourt, and L. Blanc-Féraud, "EM-MAP algorithms versus ARTUR: Theoretical and practical comparisons," in *Proc. SPIE Mathematical Methods in Medical Imaging III*, San Diego, CA, July 1994, pp. 294–304.
[26] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. MI-9, pp. 439–446, Dec. 1990.
[27] R. Mar, "Visual reconstruction with discontinuities using variational methods," *Image Vision Comput.*, vol. 10, pp. 30–38, 1992.
[28] J. L. Marroquin, "Probabilistic solution of inverse problems," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, Sept. 1985.
[29] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, Sept. 1985.
[30] E. Polak, *Computational Methods in Optimization: A Unified Approach*. NY: Academic, 1971.
[31] R. R. Schultz and R. L. Stevenson, "Stochastic modeling and estimation of multispectral image data," *IEEE Trans. Image Processing*, vol. 4, pp. 1109–1119, Aug. 1995.
[32] D. Terzopoulos, "Regularization of inverse problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 413–424, July 1986.

Pierre Charbonnier was born on September 7, 1969, in Cholet, France. He received the Ph.D. degree in electrical engineering from the University of Nice-Sophia Antipolis, France, in 1994.

He is currently in a postdoctoral position at the Telecommunications and Remote Sensing Laboratory in Louvain-La-Neuve, Belgium. His research interests include inverse problems, regularization, image reconstruction and restoration, image segmentation, deformable models, and medical imaging.

Laure Blanc-Féraud received the Ph.D. degree on image restoration from the Nice-Sophia Antipolis University, France, in 1989.

She is currently working with the CNRS at the I3S Laboratory, Sophia Antipolis. Her research interests include multiresolution image analysis, Markov random field models, nonlinear optimization, and anisotropic diffusion. Fields of application include 2-D and 3-D tomographic image reconstruction, image restoration, and motion field estimation.

Gilles Aubert was born in France on July 10, 1950. He received the These d'Etat es-science Mathématiques from Paris VI in 1986.

He is currently a professor at University Nice-Sophia Antipolis University. His research interests include calculus of variations, nonlinear partial differential equations and numerical analysis; fields of application include nonlinear elasticity and image processing.

Michel Barlaud (M'88) was born in France on November 24, 1945. He received the Doctorat d'Etat degree from the Univeristy of Paris XII.

He is currently a Professor of image processing at Ecole Superieure en Sciences Informatique, Univeristy of Nice-Sophia Antipolis. He is also Head of the Image Processing Group of the Laboratoire de Informatique Signaux et Systemes de Sophia Antipolis. His research interests are image compression using wavelets, and lattice vector quantization. His contributions to the field include biorthogonal wavelets, a method for counting lattice points on a pyramid, and a new edge adaptive quantization method.