

Fast Edge Preserving Picture Recovery by Finite Markov Random Fields

Michele Ceccarelli

Research Centre on Software Technologies-RCOST,
University of Sannio,
Via Traiano 1, 82100 Benevento, Italy

Abstract. We investigate the properties of edge preserving smoothing in the context of Finite Markov Random Fields (FMRF). Our main result follows from the definition of discontinuity adaptive potential for FMRF which imposes to penalize linearly image gradients. This is in agreement with the Total Variation based regularization approach to image recovery and analysis. We also report a fast computational algorithm exploiting the finiteness of the field, it uses integer arithmetic and a gradient descent updating procedure. Numerical results on real images and comparisons with anisotropic diffusion and half-quadratic regularization are reported.

1 Introduction

The Bayesian framework is particularly suited for solving computer vision problems as it can embed in a unique model the data consistency constraints, observation model and a priori assumptions. The underlying probabilistic model is the Markov Random Field [5], and it has been successfully applied to several inverse imaging problems such as deconvolution, denoising, interpolation, segmentation, depth estimation, shape from shading and shape from texture. The ill-posed nature of these inverse imaging problems is typically treated by recurring to Gibbs priors encompassing both the uncertainty about the solution and the desirable characteristics it should have. The generic and most popular assumption regards the smoothness of the solution [5,16]. It tends to prefer solutions characterized by local coherence and homogeneity. However, it can lead, in many situations, to over smoothed solution due to the imposition of the constraint everywhere in the image. Indeed, classical image restoration approaches are essentially based on the least squares criteria, which are basically linear and tend to smooth out edges in the output image. Therefore, the application of the smoothness constraint which preserve discontinuities has been one of the most active research areas in the computer vision community [4,11,13,14,15,18]. In particular, the concept of discontinuity adaptive prior (or edge preserving regularization) [11] is becoming even more adopted also due to the availability of fast and accurate algorithms [4,19]. Here we show that the concept of discontinuity adaptive prior can be introduced even in the context of Finite Markov Random Fields (FMRF) where the underlying space of the solution is assumed to be finite. In particular, we classify a potential function as being edge preserving if it treats in the

same way all the monotone functions in a given interval. This definition avoids to introduce the behavior of the potential function at the infinity and therefore it is more suited for FMRF. We show that in order to be edge preserving, a potential function should weight linearly the image gradient, in agreement with the recent approaches based on the Total Variation norm [3,15]. We also show how to develop a fast computational algorithm for exploiting the finiteness of the field, using integer arithmetic.

2 The MRF Approach

Here we consider the problem of restoring an image corrupted by noise. Let $I_{i,j}^0$, $i = 1, \dots, M$ and $j = 1, \dots, N$ an observed image and $I_{i,j}$ the “true” image, then our model is

$$I_{i,j}^0 = I_{i,j} + n_{i,j} \quad (1)$$

where $n_{i,j}$ denotes the noise. This problem can be solved in the context of Bayesian paradigm. The goal is to estimate the image I^* with the maximum *a posteriori* probability given I_0

$$I^* = \arg \max_I p(I|I_0). \quad (2)$$

It is well known that this MAP estimate can be solved by imposing a constrained problem [6,11,12]:

$$\arg \min_I \mathcal{R}(I) \quad \text{subject to} \quad \|I - I_0\|^2 \leq \sigma^2 \quad (3)$$

where, σ^2 is the noise variance, and $\mathcal{R}(I)$ is the prior energy functional, it measures the “quality of the image” I , in the sense that smaller values of $\mathcal{R}(I)$ correspond to “better” images. $\mathcal{R}(I)$ is the sum of local contribution from each image pixel. When there is no particular knowledge about the kind of images and the specific domain, the most natural assumption about I is its smoothness, therefore, $\mathcal{R}(I)$ should be aimed at measure the irregularities of the solution I , such irregularities being naturally depend on the derivative magnitudes of I .

Classical prior energy functionals are essentially based on the $\|\cdot\|_2$ norm of the gradient, which has the advantage of producing a set of linear equations to be satisfied by the solution. The main drawback in their use is that these functionals do not allow discontinuities in the solution, i.e. the edges are not well restored. Recently, people is even more interested in edge-preserving methods which produce much better results both from the perceptive point of view and in terms of signal-to-noise ratio. The price to be paid for these advantages is the solution of, sometimes complex, non-linear differential equation arising from the minimum condition of problem (3). In general, the prior energy has the form

$$\mathcal{R}(I) = \sum_{i,j} \phi[(D_x I)_{i,j}] + \sum_{i,j} \phi[(D_y I)_{i,j}]$$

where ϕ is the *potential function*, D_x and D_y are the discretized derivative operators in the x and y directions:

$$(D_x I)_{i,j} = (I_{i,j} - I_{i-1,j})/\delta_x \quad (D_y I)_{i,j} = (I_{i,j} - I_{i,j-1})/\delta_y.$$

In order to be a suitable potential function, ϕ should satisfy the following general assumptions:

- i) $\phi(t) \geq 0$, for any t ;
- ii) $\phi(t) = \phi(-t)$;
- iii) $\phi(t)$ is increasing for $t \geq 0$ and decreasing for $t \leq 0$.

In addition to these assumptions, a potential function ϕ is considered *edge preserving* or *discontinuity adaptive* if it further satisfies [4,11]

- iv) $\lim_{t \rightarrow \infty} \frac{\phi'(t)}{2t} = 0$;
- v) $0 \leq \lim_{t \rightarrow 0} \frac{\phi'(t)}{2t} < \infty$.

A number of edge-preserving potential functions have been proposed in literature such as: $\phi(x) = \frac{|x|^\gamma}{2(1+|x|^\gamma)}$ [7]; $\phi(x) = \log(1+t^2)$ [9]; $\phi(x) = \log(\cosh x/\gamma)$ [8] $\phi(x) = |x|$ [15]; $\sqrt{1+t^2} - 1$ [4]; $e^{-\frac{x^2}{\gamma}}$, $\frac{1}{(1+\frac{x^2}{\gamma})^2}$ [14]; and $\phi(x) = \min\{x^2, \gamma\}$ [2].

These conditions are quite natural in the context of images belonging to a continuous framework. However, in practice digital images have values over some finite finite set, such as $\{0, \dots, 255\}$. In such case the underlying image model is called *Finite Markov Random Field* (FMRF) representing the fact that $I_{i,j}$ can take only a finite set of values. In this context, the concept of infinity, of course, does not make sense, and condition iv) just represents an ideal behavior. Therefore, successfully edge preserving recovery algorithm should necessarily rely on some additional scale parameter representing thresholds which select candidate edges of the basis of gradient values which are above this threshold. In particular, the study reported in [11] classifies discontinuity adaptive potential functions in terms of the *band*, which is the interval where $\phi''(x) > 0$, outside this interval the penalty term does not depend on x , it can be either zero (no smoothing) or constant as for example the so called *line process potential function* [2]. The above potential functions are typically parametrized by a parameter γ which allow to shrink or expand the *band* of the potential thus allowing a kind of smooth threshold for the transition between uniform areas and candidate edges. This parameter being chosen as function of the image scale and the amount of edges one wants to consider inside the image.

Here we want to consider an alternative derivation of the energy potential function which does not depends on the specific values attained by each pixel and therefore is suitable for FMRF.

3 The FMRF Edge Preserving Model

In order to introduce the concept of discontinuity adaptive potential for FMRF let us consider the simple one-dimensional example plotted in Fig. 1 reporting

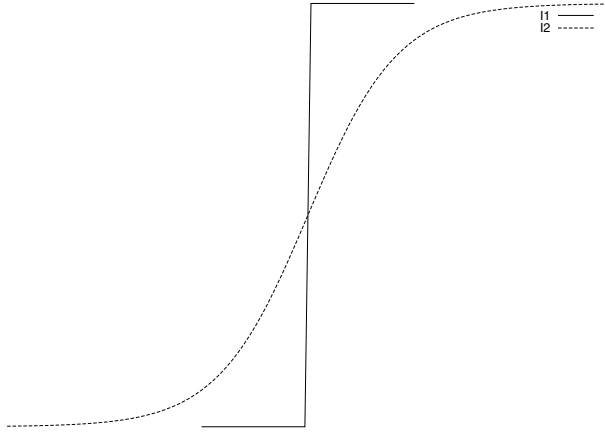


Fig. 1. The function I^1 contains an abrupt change, whereas I^2 is a smooth transition from a value to another

two functions I^1 and I^2 , the first containing an evident step discontinuity, the second being a smooth transition from a value to another. Let us consider a discretization of I^1 and I^2 as two sequences $\{I_i^1\}_{i=0,\dots,N}$ and $\{I_i^2\}_{i=0,\dots,N}$ with the same discretization step, then

$$\mathcal{R}(I^1) = \sum_{i=1}^N \phi(I_i^1 - I_{i-1}^1); \quad \mathcal{R}(I^2) = \sum_{i=1}^N \phi(I_i^2 - I_{i-1}^2).$$

If we want ϕ to be an edge preserving potential then I^1 should not be penalized more than I^2 , in the sense that the solution of the problem (3) should not be biased toward I^2 , this means that

$$\mathcal{R}(I^1) \leq \mathcal{R}(I^2).$$

This equation guarantees that sharp edges are preserved because the likelihood of solution I^1 is at least as much as that of I^2 . In other words, this model does not prefer the smooth behavior of the second solution with respect to the sharp discontinuity of the first. However, if, on the contrary, \mathcal{R} is biased toward I^1 when applied to the image I^2 , the solution of (3) will introduce artificial step discontinuities. This could be seen as the lack of the causality principle in the smoothing process aimed at solving problem (3), or equivalently that the smoothing behavior induced by such a potential can introduce artificial features during the regularization process. This last event is particular disastrous in image recovery processes where the aim is to automatic analyze image contents. Therefore we also must have

$$\mathcal{R}(I^1) \geq \mathcal{R}(I^2).$$

From these last two inequalities we derive our definition:

Definition. A potential function ϕ satisfying the general conditions (i)-(iii) is said *FRRF edge preserving* if given two monotonically increasing (decreasing) sequences $\{I_i^1\}_{i=0,\dots,N}$ and $\{I_i^2\}_{i=0,\dots,N}$ such that $I_0^1 = I_0^2$ and $I_N^1 = I_N^2$ it satisfies

$$\sum_{i=1}^N \phi(I_i^1 - I_{i-1}^1) = \sum_{i=1}^N \phi(I_i^2 - I_{i-1}^2) \tag{4}$$

Equation (4), according to the above analysis, is the right definition of edge preserving potential for FRRF. Actually it does not make use of its behavior at infinity, rather, it is based on the weight it gives to similar sequences which eventually contain abrupt changes representing edges. Neither it requires the choice, in terms of appropriate thresholds, of what a discontinuity is. It is our aim, now, to characterize the properties a function should satisfy in order to be a FRRF edge preserving potential. The first consequence of our definition is that among the potential functions listed in table 1 the Total variation norm [15] is FRRF edge preserving.

Theorem 1. *A FRRF edge preserving potential $\phi(x)$ is a linear function of x for $x \geq 0$.*

Proof. Let us consider an increasing sequence $\{f_i^1\}_{i=0,\dots,N}$, $N > 1$, and two integers ξ and η such that $0 < \xi < \eta \leq N$. Set

$$f_i^1 = \begin{cases} b & i \geq \eta \\ a & \xi \leq i < \eta \\ 0 & i < \xi \end{cases}$$

where $a, b \in \mathbf{R}$, with $a < b$. Now, let us define another increasing sequence f_2 as follows

$$f_i^2 = \begin{cases} b & i \geq \eta \\ 0 & 0 \leq i < \eta \end{cases} .$$

Both f^1 and f^2 are increasing, therefore $f_i^1 - f_{i-1}^1 \geq 0$ and $f_i^2 - f_{i-1}^2 \geq 0$, moreover $f_0^1 = f_0^2$ and $f_N^1 = f_N^2$, then let $c = b - a$, if ϕ is a FRRF edge preserving then

$$\phi(a + c) = \phi(b) = \sum_i \phi(f_i^2 - f_{i-1}^2) = \sum_i \phi(f_i^1 - f_{i-1}^1) = \phi(a) + \phi(b - a) = \phi(a) + \phi(c)$$

and this is true for any $a \geq 0$ and $c \geq 0$.

This result is, of course, not surprising. For example, most of the edge preserving functions proposed in literature have a linear behavior at infinity such as [8] [15] and [4]. In addition to the edge preserving property these three functions have the nice property of convexity, which is of help in the solution of (3). Our derivation, however, states that in order to have an edge preserving potential, this linear behavior should be always satisfied, clarifying what implicitly stated by condition (iv).

3.1 Computation of a FMRF Edge Preserving Solution

Given an FMRF edge preserving potential we want to show how we can solve problem (3) with a fast and efficient algorithm. Let us call $\mathcal{G} = \{g_0, \dots, g_{L-1}\}$ the finite set where the image pixels take values, *i.e.* $I_{i,j} \in \mathcal{G}$ and consider the maximum difference between two image values:

$$\Delta = \min_{k \neq l} |g_k - g_l|.$$

Here we develop a simple iterative algorithm aimed at the minimization of the discrete functional \mathcal{R} with the given constraint by following the iterative scheme

$$I_{i,j}^{n+1} = I_{i,j}^n + \Delta \cdot \text{sign}[\text{sign}(I_{i+1,j}^n - I_{i,j}^n) - \text{sign}(I_{i,j}^n - I_{i-1,j}^n) + \text{sign}(I_{i,j+1}^n - I_{i,j}^n) - \text{sign}(I_{i,j}^n - I_{i,j-1}^n)] \quad (5)$$

this scheme is iterated while $\|I^{n+1} - I^0\| \leq \sigma$ is true. Since $\mathcal{R}(I) \geq 0$, the following proposition states the convergence of the scheme.

Theorem 2. *The sequence of potentials $\mathcal{R}(I^n)$ generated by scheme (6) decreases monotonically.*

The proof easily follows by considering all the possible configurations (which are finite) in the neighborhood of each pixel and will be reported elsewhere. The next proposition states the causality property, which is fundamental for every iterative smoothing process. Roughly speaking, the causality principle states that each feature at a coarse scale must have a cause at a finer scale. This means that the smoothing process does not introduce spurious features. Formally, it can be shown that every causal smoothing process must be governed by, or be the discretized version of, a parabolic partial differential equation obeying a maximum principle [1].

Theorem 3. *The scheme (6) satisfies*

$$\min\{I_{i,j}^n, I_{i-1,j}^n, I_{i+1,j}^n, I_{i,j-1}^n, I_{i,j+1}^n\} \leq I_{i,j}^{n+1} \leq \max\{I_{i,j}^n, I_{i-1,j}^n, I_{i+1,j}^n, I_{i,j-1}^n, I_{i,j+1}^n\}$$

Proof. Let $m = \min\{I_{i,j}^n, I_{i-1,j}^n, I_{i+1,j}^n, I_{i,j-1}^n, I_{i,j+1}^n\}$, and $M = \max\{I_{i,j}^n, I_{i-1,j}^n, I_{i+1,j}^n, I_{i,j-1}^n, I_{i,j+1}^n\}$ if $I_{i,j}^n = m$ then it is easy to verify that

$$[\text{sign}(I_{i+1,j}^n - I_{i,j}^n) - \text{sign}(I_{i,j}^n - I_{i-1,j}^n) + \text{sign}(I_{i,j+1}^n - I_{i,j}^n) - \text{sign}(I_{i,j}^n - I_{i,j-1}^n)] \leq 4$$

and, in this case, from (6)

$$I_{i,j}^{n+1} = I_{i,j}^n + \Delta \geq m.$$

In the other cases, from the definition of Δ , we have

$$I_{i,j}^n \geq m + \Delta$$

and therefore

$$I_{i,j}^{n+1} \geq I_{i,j}^n - \Delta \geq m$$

where the first inequality comes from (6). The proof that $I_{i,j}^{n+1} \leq M$ is analogous.

4 Experiments and Comparisons

In this section we will present some experimental result about the application of the developed algorithm to synthetic and real grayscale. In particular the algorithm reads as:

Digital Picture Recovery Algorithm

//Input: a discrete image $I_{i,j}^0, i = 1, \dots, M$ and $j = 1, \dots, N$
 //Output: the recovered image

1. Estimate $\tilde{\sigma}$
2. $I_{i,j} := I_{i,j}^0$
3. $n := 0$;
4. **while** $(\sum_{i,j} (I_{i,j}^n - I_{i,j}^0)^2 \leq MN \cdot \sigma)$
5. Apply (6) to each image pixel
6. $n := n + 1$
7. **end while**
8. output I^n .

Note that there are several methods to perform step 1. Our implementation adopts a variant of the method proposed in [10]. In particular,

$$\tilde{\sigma} = \frac{1}{36} \text{Variance}(I^0 \otimes \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{bmatrix}) \quad (6)$$

where \otimes represents the convolution operator. Note that the algorithm does not require any parameter. In our implementation we have the choice to implement filter (6) in a recursive manner, *i.e.* the updating is performed in place, this kind of updating produces a significant speed-up of the convergence while maintaining the causality properties of the method. In any case the experiments presented below are based on batch updating. Since we also report computing time, the adopted computing platform is significant, all the experiments were performed on a 600 MHz Pentium II Linux Personal Computer.

In order to evaluate the behavior of the algorithm, and to compare it with other edge preserving denoising, we artificially add to the original image some amount of noise and then measure the quality of the reconstruction as function of the iteration. Here we compare the algorithms with well known edge preserving image recovery techniques such as the half-quadratic regularization by the ARFUR algorithm [4,6] and anisotropic diffusion [3,14,15]. It is well known that the quantitative measures of image reconstruction may often fail with respect to perceptually plausible measures. For example the mean squared error measure tends to compress small errors and to overweight large errors. In this paper we adopt as a quantitative measure of the reconstruction the so called *Mean Error* (ME) defined as

$$ME(I, I^0) = \frac{1}{n} \sum_{i,j} |I_{i,j} - I_{i,j}^0| \quad (7)$$

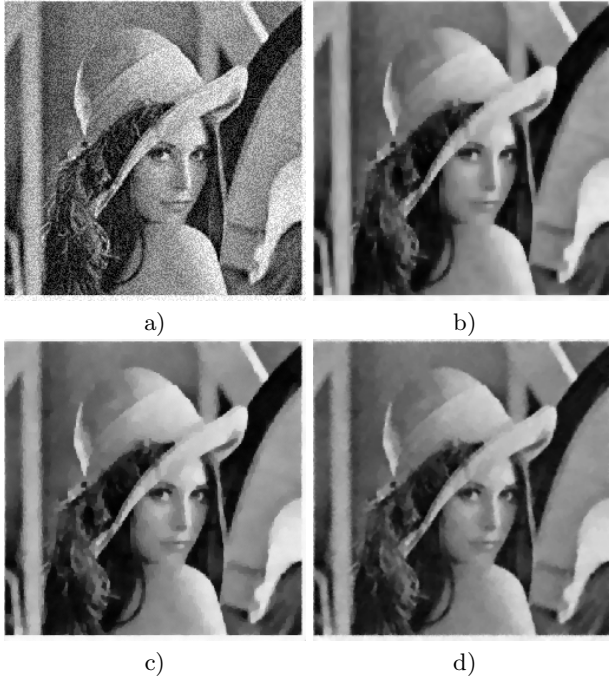


Fig. 2. The lena image, corrupted by uniform noise at 8.5 dB of SNR a), and its reconstruction by the proposed method b), the anisotropic diffusion c) and half quadratic regularization d) ($\alpha = 0.075$)

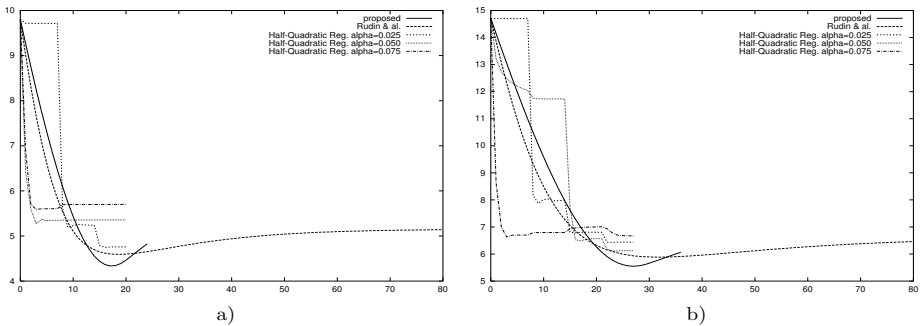


Fig. 3. The ME measure as function of the iteration for the reported algorithms. For the half quadratic regularization the measure is computed at each iteration of the iterative algorithm adopted to solve the inner linear system, which our case is a conjugate gradient algorithm with a multigrid preconditioner. This figure refers to uniform noise at 14.5 dB a) and at 11 db b) of SNR.

where n is the number of image pixels. The first case adopts uniformly distributed additive noise. In particular we add uniform noise at an amount of 8.5dB, 11dB and 14.5dB of Signal to Noise Ratio (SNR). The corrupted image and the corresponding reconstruction are reported in Fig. 2. As it can be seen from the images there is no significant difference between the reconstructions, at least from the perceptual level. In order to quantitatively appreciate the behavior of the algorithm we report the ME as function of the iteration number. As Fig. 3 shows, the proposed algorithm compares well in terms of quality of reconstruction with the other algorithms reported. For what concerns the parameters adopted for the generation of such figure let us mention that our algorithm does not need any free parameters, whereas for the case of anisotropic diffusion we choose the time step as 0.5, the maximum number of iteration is 80 and the noise variance the same that estimated by (6). Whereas, for the half-quadratic reconstruction we adopted the regularization parameters reported in the figure, and fixed the maximum number of outer iterations to four and the maximum number of inner iteration of the preconditioned conjugate gradient algorithm to seven. The above figures also show that the ARTUR algorithm has a very fast convergence rate however one should consider that the price in terms of computation is much higher with respect to the proposed algorithm. Specifically, for the reported 256×256 grayscale image the computing times over the adopted platform of each inner iteration, are 0.027, 0.076 and 0.562 seconds respectively for the proposed algorithm, an isotropic diffusion and half-quadratic regularization. This means that each inner iteration takes 5% of the the time of half quadratic algorithm, and 14% of the time of each non linear diffusion iteration. If, in addition, we consider that the right regularization parameter must be typically chosen in an experimental trial and error fashion, the advantage of the proposed method is even more evident.

5 Conclusions

We have reported an image recovery algorithm which is based on the Finite Markov Random Field model. We have investigated the properties of edge preserving potential functions for FMRF and clarified that the linear behavior of potential functions is fundamental for convex edge preserving priors. The resulting algorithm is fast and efficient, does not require any choice of free parameters.

References

1. L. Alvarez, P. L. Lions, F. Guichard, and J. M. Morel, "Axioms and Fundamental equations of Image Processing", *Archives for Rational Mechanics and Analysis*, vol. 16(9), pp. 200-257, 1993
2. A. Blake, and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MASS., 1987.
3. M. Ceccarelli, V. De Simone, "Well Posed Anisotropic Diffusion for Image Denoising", em IEE Proceedings Proceedings-Vision, Image and Signal Processing, vol 149, 4, pp. 244-252, 2002.

4. P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic Edge-Preserving Regularization in Computed Imaging", *IEEE Transactions on Image Processing*, vol. 5, pp. 298-311, 1997.
5. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.
6. S. Geman and G. Reynolds, "Constrained Restoration and the Recovery of Discontinuities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 367-383, 1992.
7. S. Geman and D. E. McClure, "Bayesian image analysis: an application to single photon emission tomography", in *Proc. Statistical Computational Section, Amer. Statistical Assoc.*, Washington, DC, 1995, pp. 12-18.
8. P. J. Green, "Bayesian reconstruction for emission tomography using a modified EM algorithm", *IEEE Transactions on Medical Imaging*, vol. 9, pp. 84-93, 1990.
9. T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data using Gibbs Priors", *IEEE Transactions on Medical Imaging*, vol. 8, pp. 194-202, 1990.
10. J. Immerkaer, "Fast Noise Variance Estimation", *CVGIP: Image Understanding*, vol. 64, pp. 300-302, 1996.
11. S. Z. Li, "On Discontinuity-Adaptive Smoothness Priors in Computer Vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 576-586, 1995.
12. S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer, Berlin, 1995.
13. D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems", *Communications on Pure and Applied Mathematics*, vol. 42, pp. 577-685, 1989.
14. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 345-362, 1990.
15. L. Rudin, S. Osher and E. Fatemi "Nonlinear total variation noise removal algorithms", *Physica D*, vol. 60, pp. 259-268, 1992.
16. T. Poggio, V. Torre, C. Koch, "Computational Vision and Regularization Theory", *Nature*, vol. 317, pp. 314-319, 1985.
17. Y. L. You, W. Xu, A. Tannebaum and M. Kaveh, 'Behavioral Analysis of Anisotropic Diffusion in Image Processing', *IEEE Transactions of Image Processing*, vol. 5, pp. 1539-1553, 1996.
18. D. T. Terzopoulos, "Regularization of inverse visual problems involving discontinuities", *IEEE Transactions on PAMI*, vol. 8(4), pp. 413-242, 1986I.
19. C. R. Vogel and E. Oman, 'Iterative Methods for Total Variation Denoising', *SIAM Journ. on Scientific Computing*, vol 17, pp. 227-238, 1996.