A Variational Model for P+XS Image Fusion

C. Ballester, V. Caselles, J. Verdera
University Pompeu Fabra
Departament de Tecnologia
Pg. Circumvalació 8, 08003 Barcelona, Spain
e-mail: vicent.caselles@upf.edu

B. Rougé
CNES
Av. Ed Belin, 31401 Toulouse, France
e-mail: Bernard.Rouge@cnes.fr

Abstract

We propose an algorithm to increase the resolution of multispectral satellite images knowing the panchromatic image at high resolution and the spectral channels at lower resolution. Our algorithm is based on the assumption that, to a large extent, the geometry of the spectral channels is contained in the topographic map of its panchromatic image. This assumption, together with the relation relating the panchromatic image to the spectral channels, and the expression of the low resolution pixel in terms of the high resolution pixels given by some convolution kernel followed by subsampling, constitute the elements to construct an energy functional (with several variants) whose minima will give the reconstructed spectral images at higher resolution. We shall discuss the well foundedness of the above approach and describe our numerical approach. Finally, some experiments are displayed.

1 Introduction

A grey level image can be realistically modeled as a real function $u(x)$ where x represents an arbitrary point of a rectangle Ω in $\mathbb{R}^2$ and $u(x)$ denotes the grey level at x. Typically $u(x)$ represents the photonic flux over a wide band of wavelengths and we have a proper grey level image. Below, we shall refer to this image as the panchromatic image. A multispectral image may be represented by a function $\vec{u}$ from $\mathbb{R}^2$ to $\mathbb{R}^m$ where m represents the number of spectral channels. For colour images, typically, $m = 3$ if we consider the usual R, G, B channels. If we add the infrared channel to the colour channels we have a multispectral image with $m = 4$. In this case each coordinate of $\vec{u}(x)$ represents the intensity corresponding to a spectral channel, when the photonic flux is subjected to an spectral selective filter, be it in the visible range, the infrared or the ultraviolet one. We shall say that the panchromatic image u corresponds to the multispectral image $\vec{u}$ if $u(x)$ has been obtained by adding (with some mixing coefficients) the coordinates of $\vec{u}$ which represent the energies of different spectral bands. In other words, if $\vec{u} = (u_1, ..., u_m)$ and $\alpha_1, ..., \alpha_m$ represent the mixing coefficients to compute the total spectral energy from the above channels, the corresponding panchromatic image will be given by

$$u(x) = \alpha_1 u_1(x) + ... + \alpha_m u_m(x). \quad (1)$$

Let us mention that in digital images, the only accessible information is a sampled and quantized version of u, $u(i, j)$, where (i, j) is a set of discrete points (in general on a grid) and $u(i, j)$ belongs in fact to a discrete set of values, $0, 1, ..., 255$ in many cases. Since, by Shannon's theory, we can assume that $u(x)$ is recoverable at any point from the samples $u(i, j)$, as a first approximation, we may assume that the image $u(x)$ is known in a continuous domain, up to the quantization noise.

The purpose of this paper will be to describe a method to increase the resolution of satellite multispectral images when we know the corresponding panchromatic image at a higher resolution. We shall assume that the panchromatic image u has been sampled at higher resolution, and we have an image $u(i, j)$ of size $N \times N$ pixels. The spectral channels have been sampled at a lower resolution giving images of size $\frac{N}{p} \times \frac{N}{p}$ (typically $p = 2$, or 4) which we shall denote by $\vec{u}^p = (u_1^p, ..., u_m^p)$, the superscript p being explicitly included to stress the loss of resolution of the multispectral data. Our purpose will be to reconstruct the high resolution multispectral image $\vec{u} = (u_1, ..., u_m)$, which will be an image of size $N \times N$, knowing the data u and $\vec{u}^p$. For that we shall take into account several constraints imposed by the data generation model. In particular, the low resolution pixel is formed from the high resolution pixel by a low pass filtering followed by a subsampling. If we denote by k the impulse response of this filter, we may write

$$u_i^p(i, j) = k * u_i(i, j) \quad \forall i, j \in \{0, ..., \frac{N}{p} - 1\}. \quad (2)$$

The precise kernel k used in our simulations will be specified below. Obviously, we have to respect also the relation (1). Finally, we shall use the geometric information contained in the panchromatic image. Indeed, we shall constraint the geometry of the spectral channels at higher resolution to follow the geometry of the panchromatic image. This constitutes the main feature of our approach and needs further explanation. First, we shall explain what do we understand by the geometric information contained in the panchromatic image, and more generally, in any scalar image, including any

spectral channel. Then we shall explain the reasons which support the underlying assumption that *images taken on different spectral bands share common geometric information*. For that we shall review the main conclusions in [7],[9].

Before going into the details of our method, let us say that there exists a huge literature on the subject. The most performing methods are based on the injection of high frequency components (corresponding to spatial details present in the high resolution panchromatic image) in interpolated versions of the multispectral data [12], or improvements based on multiresolution analysis of P+XS images [1]. For more information we refer to [1],[12], and references therein.

1.1 Mathematical morphology of scalar images

In this subsection, we shall consider scalar images, that is, images with a single channel, be it colour (or any other spectral channel) or grey level.

What is the geometric information content of an image ? This is the question we want to discuss briefly here. In this paragraph, we are simply summarizing some arguments contained explicitly or implicitly in the Mathematical Morphology theory [11], which were further developed in [6].

The sensors of a camera or a CCD array transform the continuum of light energies to a finite interval of values by means of a nonlinear contrast function g. The contrast change g depends on the properties of the sensors, but also on the illumination conditions and the reflection properties of the objects, and those conditions are generally unknown. Images are observed modulo an arbitrary and unknown contrast change. These observations lead the physicist and psychologist M. Wertheimer [13] to state as a basic principle that the gray level is not an observable.

Mathematical Morphology recognized contrast invariance as a basic invariance requirement and proposed that image analysis operations take into account this invariance principle [11, 8]. With this principle, an image u is a representative of an equivalence class of images v obtained from u via a contrast change, i.e., $v = g(u)$ where g, for simplicity, will be a continuous strictly increasing function. Under this assumption, an image is characterized by its upper (or lower) level sets $X_\lambda = [u \geq \lambda] = \{x : u(x) \geq \lambda\}$ (resp. $X'_\lambda = [u \leq \lambda] = \{x : u(x) \geq \lambda\}$). Moreover, the image can be recovered from its level sets by the reconstruction formula

$$u(x) = \sup\{\lambda : x \in X_\lambda\}.$$

Thus, according to the Mathematical Morphology doctrine, the reliable information in the image is contained in the level sets, independently of their actual levels. Thus, we are led to consider that the geometric information, the shape information, is contained in those level sets.

We can further describe the level sets by their boundaries, $\partial X_\lambda u$, which are, under suitable very general assumptions, Jordan curves. Jordan curves are continuous maps from the circle into the plane $\mathbb{R}^2$ without crossing points. Indeed, in [2] it is proved that if u is a function whose upper level sets $X_\lambda u$ are sets of finite perimeter (in particular, if u is a function of bounded variation [3]), then the boundaries of level sets can be described by a countable family of Jordan curves with finite length. The family of all level lines of an image was called the *topographic map* [6]. We can conceive the topographic map as a tool giving a complete description of the geometry for grey level images.

Still, we have to go a step further in the description of the geometry of the topographic map. We shall assume that that we can compute the normal unit vector to the level lines of u, i.e., there is a vector field $\theta : \Omega \to \mathbb{R}^2$ with $|\theta(x)| \leq 1$ a.e. such that $\theta \cdot \nabla u = |\nabla u|$. To give a rigorous sense to this expression, we could assume for instance that u is a function of bounded variation in Ω, and we have that (almost all) its level sets $[u \geq \lambda]$ are sets of finite perimeter whose boundaries are Jordan curves, and there is a vector field θ which coincides with the unit normal vector at almost any point of them (see Section 5). Then the identity $\theta \cdot \nabla u = |\nabla u|$ can be understood as equality of two measures. In practice, at the discrete level, θ can be defined by the relation $\theta(x, y) = \frac{\nabla u(x,y)}{|\nabla u(x,y)|}$ when $\nabla u(x, y) \neq 0$, and $\theta(x, y) = 0$ when $\nabla u(x, y) = 0$. This vector field will be the right analytic tool which we require to impose the constraint that the geometry of any function v is specified by the geometry of u, indeed, by θ.

1.2 Geometry and color in natural images

What is geometric content of a color image ? Obviously, the answer to this question is quite complex and, strictly speaking, it cannot be reduced to the geometry of its associated intensity image. Indeed, counterexamples can be given were color objects exist with a constant intensity. But what happens in images of natural scenes ? Will the light create color patterns with color edges with a constant intensity ? This problem was addressed in [7] and the authors experimentally checked the hypothesis that the essential geometric contents of a color image is contained in the level lines of the corresponding intensity image. Indeed, they designed an algorithm to constrain the color channels of a given image to have the same geometry (i.e. the same level lines) as the grey level [7]. The algorithm can be briefly described: *replace the colors in an image by their conditional expectation with respect to the grey levels*. If the hypothesis above is sound, then this algorithm should not alter the colors of the image or its visual aspect. We refer to [7] for this experimental discussion.

1.3 The case of satellite multispectral images

In the case of multispectral satellite images, the analogous assumption would be that images of the same scene taken on different wavelength bands would share a common geometric information. Indeed, this assumption was experimentally studied in [9]. The experiments were done on images

of four spectral channels corresponding to the blue, green, red and near infrared regions. Two channels were compared by means of their topographic maps. Two types of comparison were proposed: by means of the unit normal vector field of the topographic map, and by means of pieces of its level lines [9]. With both comparison procedures the conclusion was the same, the channels which are in the visible region share a large portion of its topographic map, this amount decreases, but it is still large for the infrared channel. For the red and infrared images, it was also showed that, after contrast inversion, there is still a portion of the topographic map which is common to the topographic map of the blue channel [9].

Thus, based on the above arguments, we shall adopt the hypothesis that for satellite multispectral images, *to a large extent, the geometry of the spectral channels is contained in the topographic map of its panchromatic image.* This assumption, together with (1) and (2) is the basis of our variational approach to the problem studied here.

Let us explain the plan of the paper. Section 2 is devoted to describe the energy functional of our variational model. Finally, Section 3 is devoted to the description of the algorithm and the numerical experiments. We end up in Section 4 with some conclusions. Finally, in Section 5 we briefly comment on the mathematical justification of the problem.

2 The energy functional: continuous an discrete description

2.1 The continuous formulation

To fix ideas, we assume that the multispectral image is given by a function $\vec{u} : \Omega \to \mathbb{R}^3$ where Ω is a rectangle of $\mathbb{R}^2$, say $[0,1]^2$. We shall denote the coordinates of $\vec{u}$ by (X_1, X_2, X_3) and call them the red, green and blue channels. As above, we denote by u the intensity image corresponding to $\vec{u}$.

Assume that we are given the image u on Ω and that we know the values of $\vec{u}$ on a sampling grid $S \subseteq \Omega$ whose points will be called the low resolution pixels. Let us denote by $\vec{u}^S$ the known values of $\vec{u}$ on S, in coordinates, $\vec{u}^S = (X_1^S, X_2^S, X_3^S)$. Recall that the low resolution pixel is formed from the high resolution pixel by a low pass filtering followed by subsampling. Let k be the impulse response of this filter. Our purpose will be to reconstruct u from the data u and $\vec{u}^S$. According to the discussion of Section 1, we should impose the following two relations

$$u(x) = \alpha_1 X_1(x) + \alpha_2 X_2(x) + \alpha_3 X_3(x) \qquad (3)$$

where $\alpha_1, \alpha_2, \alpha_3 > 0$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$, are the coefficients which give the intensity image in terms of the spectral channels, and

$$X_n^S(i,j) = k * X_n(i,j), \forall (i,j) \in S,\ n = 1,2,3 \qquad (4)$$

which correspond to (1), and (2), respectively. To give a sense to the relations in (4), we need to assume that it is possible to evaluate $k * X_n$ at any point of S. For that, we shall assume that

(H) k is the kernel of a convolution operator mapping $L^2(\Omega)$ into $C(\overline{\Omega})$.

Under assumption (H), for any point $(i,j) \in \Omega$, the map which to any $f \in L^2(\Omega)$ associates the value

$$k * f(i,j) = \int_\Omega k((i,j) - (x,y)) f(x,y)\, dx\, dy$$

is a continuous linear functional in $L^2(\Omega)$ and identities (4) have a sense.

The problem of recovering (X_1, X_2, X_3) from u and (X_1^S, X_2^S, X_3^S) is ill-posed. Indeed, conditions (3) and (4) do not determine uniquely the vector (X_1, X_2, X_3), and the problem involves the inversion of a convolution equation. Typically one is led to a regularization method. The geometric requirement that the geometry of the images X_1, X_2, X_3 is given by the geometry of the intensity u will give the required regularization.

Constraining the geometry of the spectral channels. Since θ has the direction of the normal to the level lines of u, the counterclockwise rotation of angle $\pi/2$, denoted by $\theta^\perp$, represents the tangent vector to the level lines of u. In this case, if the spectral channels share the geometry of the panchromatic image, we have

$$\theta^\perp \cdot \nabla X_n = 0, \quad n = 1,2,3. \qquad (5)$$

Obviously, the relations (5) cannot be exactly satisfied and we have to impose them in a variational framework (together with the other constraints discussed at the introduction) by minimizing the sum of integrals

$$\sum_{n=1}^{3} \gamma_n \int_\Omega |\theta^\perp \cdot \nabla X_n|^p \qquad (p = 1,2), \qquad (6)$$

where $\gamma_n > 0$. The same relationship can be imposed in a slightly different way. Indeed, given the vector field θ of unit normals to the level sets of u, and assuming that $X_1, X_2, X_3 \in BV(\Omega)$ we shall require that $|\nabla X_n| = \theta \cdot \nabla X_n$, $n = 1,2,3$. Again, these relations can be imposed in a variational framework by minimizing the sum of integrals

$$\sum_{n=1}^{3} \gamma_n \int_\Omega (|\nabla X_n| - \theta \cdot \nabla X_n). \qquad (7)$$

Let us write functional (7) in a convenient form for computational purposes. For that, we integrate by parts the second term of each integral, and using $\theta \cdot \nu = 0$ on $\partial\Omega$ (where ν is the outer unit normal to $\partial\Omega$), we may write (7) in the form

$$\sum_{n=1}^{3} \gamma_n \int_\Omega (|\nabla X_n| + div\, \theta \cdot X_n). \qquad (8)$$

In both cases, the constants γ_n ($n = 1, 2, 3$) permit to control the relative weight assigned to each channel.

Imposing (3) and (4) in a variational framework. We impose the constraint (3) by minimizing the integral term

$$\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 - u)^2 \tag{9}$$

We may impose (4) by minimizing the sums

$$\sum_{(i,j)\in S} (k * X_n(i,j) - X_n^S(i,j))^2. \tag{10}$$

We may write the above relations in an integral form, and this will be useful in order to write the Euler-Lagrange equations in a more compact form. We need some notation for that. Let $\delta_{(i,j)}$ be the Dirac's delta at the point (i,j). Let $\Pi_S = \sum_{(i,j)\in S} \delta_{(i,j)}$ be the Dirac's comb defined by the grid S. Then, we may write (10) as the integral

$$\int_\Omega \Pi_S (k * X_n(x,y) - X_n^S(x,y))^2 \, dx \, dy \tag{11}$$

where $X_n^S(x,y)$ denotes an arbitrary extension (from S to Ω) of $X_n^S(i,j)$ as a continuous function. Since the integrand of (11) is multiplied by Π_S, the term (11) does not depend on the particular extension of X_n^S, $n = 1, 2, 3$.

We shall impose a further constraint on R, G, B. Indeed, let $M_n = \max_{(i,j)\in S} \max(\frac{u(i,j)}{\alpha_n}, X_n^S(i,j))$, $n = 1, 2, 3$. Then we shall impose that

$$0 \le X_n \le M_n, \quad n = 1, 2, 3. \tag{12}$$

These constraints are useful for a mathematical justification of the algorithm.

The energy functional. Thus, we propose to compute the high resolution multispectral images X_1, X_2, X_3 by minimizing the energy functional:

$$\sum_{n=1}^3 \gamma_n \int_\Omega |\theta^\perp \cdot \nabla X_n|^p + \lambda \int_\Omega (\sum_{n=1}^3 \alpha_n X_n - u)^2 +$$

$$\mu \sum_{n=1}^3 \int_\Omega \Pi_S \Big((k * X_n(x,y) - X_n^S(x,y))^2 \tag{13}$$

subject to $0 \le X_n \le M_n$, $n = 1, 2, 3$.

For the purposes of comparison, let us also write a variant of (13) which is based on (8)

$$\sum_{n=1}^3 \gamma_n \int_\Omega (|\nabla X_n| + div\, \theta \cdot \nabla X_n) +$$

$$\lambda \int_\Omega (\sum_{n=1}^3 \alpha_n X_n - u)^2 + \tag{14}$$

$$\mu \sum_{n=1}^3 \int_\Omega \Pi_S \Big((k * X_n(x,y) - X_n^S(x,y))^2$$

Figure 1: The color reference image $\vec{u}^{ref}$.

subject to the same constraints as (13). Experiments on this functional have been reported in [4]. Finally, let us note that the first integral in (14) was also used in [5] in the context of filling-in by joint interpolation of vector fields and gray levels, in order to constrain the vector field θ and the image u to be related by $\theta \cdot \nabla u = |\nabla u|$, but, in this case, both θ and u where unknown.

In both functionals we take $\gamma_1, \gamma_2, \gamma_3, \lambda, \mu > 0$ (in practice, all these parameters are taken equal to 1), and $p = 1$, or 2. Notice that functional (13) is invariant under the change $\theta \to -\theta$, or, in other words, it is contrast invariant. The main advantage of functional (13) is that the corresponding Euler-Lagrange equations when $p = 2$ are linear and the steepest descent method converges to a minimum in a much faster way than (14).

2.2 The discrete formulation

To proceed with the discrete numerical algorithm, we assume that the panchromatic image u is given on $\{0, 1, ..., N-1\} \times \{0, ..., N-1\}$. We replace the gradients in (14) by its discrete approximation: for any scalar function f we shall use the notation $\nabla^{+,+} f = (\nabla_x^+ f, \nabla_y^+ f)$, $\nabla^{+,-} f = (\nabla_x^+ f, \nabla_y^- f)$, $\nabla^{-,+} f = (\nabla_x^- f, \nabla_y^+ f)$, $\nabla^{-,-} f = (\nabla_x^- f, \nabla_y^- f)$ where $\nabla_x^+ f(i,j) = f(i+1,j) - f(i,j)$, $\nabla_x^- f(i,j) = f(i,j) - f(i-1,j)$, and similar expressions hold for $\nabla_y^{\pm,\pm} f(i,j)$. We use the notation $\theta^{\alpha,\beta} = \frac{\nabla^{\alpha,\beta} u}{|\nabla^{\alpha,\beta} u|}$ if $\nabla^{\alpha,\beta} u \neq 0$ and $\theta^{\alpha,\beta} = 0$ if $\nabla^{\alpha,\beta} u = 0$.

For simplicity, we shall only describe the discrete formu-

lation of (13) which can be written as

$$\sum_{n=1}^{3} \frac{\gamma_n}{4} \sum_{\alpha,\beta=+,-} \sum_{i,j=0}^{N-1} |\theta^{\alpha,\beta}(i,j)^{\perp} \cdot \nabla^{\alpha,\beta} X_n(i,j)|^2 +$$

$$\lambda \sum_{i,j=0}^{N-1} (\sum_{n=1}^{3} \alpha_n X_n(i,j) - u(i,j))^2 +$$

$$\mu \sum_{n=1}^{3} \sum_{(i,j)\in S} \Big((k * X_n(i,j) - X_n^S(i,j))^2,$$

$$\text{subject to } 0 \leq X_n \leq M_n,\ n = 1,2,3, \tag{15}$$

Observe that we have used simultaneously the four discretizations for the gradient, since using only one of them may produce some artifacts or asymmetries in the results. It is still to be precised the kernel k and the discrete approximation of $k * \{X_1, X_2, X_3\}$. This will be done in next Section.

3 Algorithm and numerical experiments

Figure 2: The panchromatic image at resolution 0.8 m/pixel, denoted in the text by u.

To minimize (15) using the gradient descent method, we iteratively actualize the solution using the equation

$$\begin{aligned} X_n^{p+1} &= X_n^p + \\ &+\Delta t \tfrac{\gamma_n}{4} \textstyle\sum_{\alpha,\beta=+,-} \operatorname{div}^{\alpha*,\beta*} \left(< \theta^{\alpha,\beta}, \nabla^{\alpha,\beta} X_n^p > \theta^{\alpha,\beta} \right) \\ &-\mu \Delta t k^t * \left(\Pi_S (k * X_n^p - X_n^S) \right) \\ &-\lambda \alpha_n \Delta t (\textstyle\sum_{n=1}^{3} \alpha_n X_n^p - u). \end{aligned} \tag{16}$$

The constraint that X_n, $n = 1,2,3$, should remain in the range between 0 and an upper value can be imposed after each iteration by brute force. To avoid a cumbersome expression we have avoided to write in detail the terms involving k.

3.1 Description of the data

To test our method we shall dispose of some reference images which we have been constructed by the CNES for this purpose. Thus, we are given a reference multispectral image $\vec{u}^{ref} = (R_{ref}(i,j), G_{ref}(i,j), B_{ref}(i,j))$, $i,j \in \{0,...,N-1\}$, which has a resolution of $0.8\ m/pixel$ and satisfies Shannon's rule. We shall consider the case where $\vec{u}^{ref}$ are the usual color red, green, and blue channels, but they could also represent the near infrared, red and green channels. The images take integer values in the range between 0 and 1023. Instead of X_1, X_2, X_3 we shall use the usual notation R, G, B. The corresponding intensity image is computed by

$$u^{ref}(i,j) = \alpha_R R_{ref}(i,j) + \alpha_G G_{ref}(i,j) + \alpha_B B_{ref}(i,j).$$

To construct an image at resolution $1.6\ m/pixel$ we subsample the image $\vec{u}^{ref}$ by a factor of 2. Let us denote this image by $\vec{u}^{1.6} = (R_{1.6}, G_{1.6}, B_{1.6})$. Then

$$\vec{u}^{1.6}(i,j) = \vec{u}^{ref}(2i,2j) \quad i,j \in \{0,...,\frac{N}{2}-1\}. \tag{17}$$

To construct an image at resolution $3.2\ m/pixel$ either we average the pixels of $\vec{u}^{1.6}$ with a window of size 2 followed by a subsampling of factor 2 (or we average the pixels of $\vec{u}^{ref}$ and subsample them by a factor of 4). Let us denote this image by $\vec{u}^{3.2} = (R_{3.2}, G_{3.2}, B_{3.2})$. The precise relation is

$$\begin{aligned} \vec{u}^{3.2}(i,j) &= \tfrac{1}{4} \textstyle\sum_{(l,m)\in\{0,1\}} \vec{u}^{1.6}(2i+l, 2j+m) \\ &= \tfrac{1}{4} \textstyle\sum_{(l,m)\in\{0,2\}} \vec{u}^{ref}(4i+l, 4j+m). \end{aligned} \tag{18}$$

for $i,j \in \{0,...,\frac{N}{4}-1\}$. Our data will be constituted by u^{ref} and $(R_{1.6}, G_{1.6}, B_{1.6})$, or by u^{ref} and $(R_{3.2}, G_{3.2}, B_{3.2})$, depending if we want to test our algorithm to increase the resolution by a factor 2 or 4. The particular kernel is described either by the relation between $\vec{u}^{1.6}$ and $\vec{u}^{ref}$ given in (17), or by the relation between $\vec{u}^{3.2}$ and $\vec{u}^{ref}$ given in (18). These kernels were suggested to us by the CNES with the purpose of simple numerical experimentation.

3.2 Numerical experiments

We have tested both cases: an increase of resolution by a factor 2, and 4. To simplify our presentation we shall concentrate on the case of factor 4.

As we have said in the previous subsection, our channels X_1, X_2, X_3, represent the usual R, G, B channels. In our experiments, we have taken $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$ (but they could be any given values), and $\gamma_1 = \gamma_2 = \gamma_3 = 1$, $\lambda = \mu = 1$. Let us mention that, in our experiments below, we have

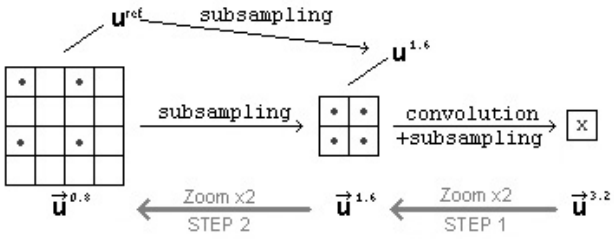

Figure 3: Our two-step strategy.

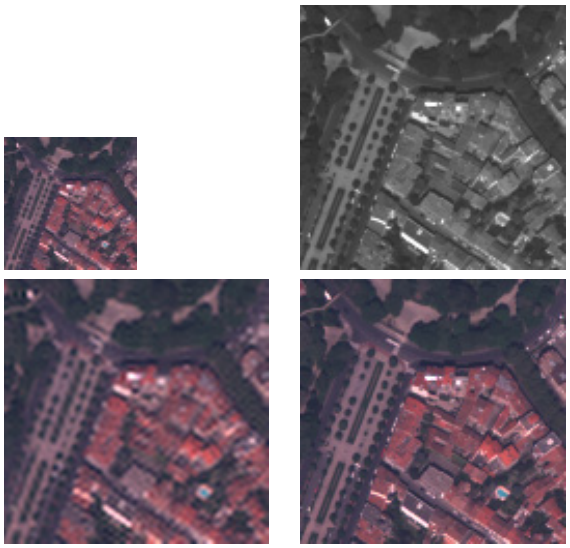

Figure 4: a) Top left: color image at resolution 3.2 m/pixel, $\vec{u}^{3.2}$. b) Top right: the panchromatic image obtained by subsampling u^{ref} by a factor 2 (resolution 1.6 m/pixel. c) Bottom left: initialization of (16) obtained by a replication of $\vec{u}^{3.2}$ by a factor 2. d) Top right: the result $\vec{u}^{inter}$ obtained using (16).

checked that the functional attains its minimum, in the sense that the value of each term of the functional (15) is near to 0.

The reference image $\vec{u}^{ref} = (R_{ref}, G_{ref}, B_{ref})$ is displayed in Figure 1. Figure 2 displays the corresponding panchromatic image u^{ref}. Both are at resolution 0.8 m/pixel.

The reference image permits us to assess the quality of the reconstructed image. We shall compare the reconstructed images (R, G, B) with the reference images $(R_{ref}, G_{ref}, B_{ref})$ both visually and by displaying some error measures. We shall compute the PSNR, the maximum, the ℓ^1 and ℓ^2 errors on the whole image and on certain representative regions. We also display the histogram of errors, and we make explicit some percentiles.

We minimize (15) using the gradient descent equations given in (16). To accelerate its convergence we decompose the algorithm in two steps (see Figure 3). In the first step we compute a zoom of factor 2 of $\vec{u}^{3.2} = (R_{3.2}, G_{3.2}, B_{3.2})$ (see Figure 4.a). For that, we need a panchromatic image, let us call it $u^{1.6}$, at the resolution $1.6m/pixel$ which we create by subsampling u^{ref} by a factor of 2 (see Figure 4.b). Using this panchromatic image $u^{1.6}$ (which also permits us to compute the vector field θ entering in (15)), and the color image $\vec{u}^{3.2}$, we minimize (15) by means of the gradient descent equations given in (16). The kernel k is given by the first of equations in (18). As initial conditions for (16) we simply take a replication by a factor 2 of $\vec{u}^{3.2}$, see Figure 4.c. We obtain as a result an intermediate image $\vec{u}^{inter}$ at a resolution of 1.6 m/pixel. This intermediate result is displayed in Figure 4.d.

In the second step we compute a zoom of factor 2 of the result $\vec{u}^{inter}$ of the previous step reaching finally the resolution of $3.2m/pixel$. This time the panchromatic image is u^{ref} (resolution $3.2m/pixel$), and the color image at the resolution $1.6m/pixel$ is $\vec{u}^{inter}$. Again, we minimize (15) (with convolution operator $k = \delta$) by means of the gradient descent equations (16). We take as initial condition the image obtained by a simple replication of the pixels of $\vec{u}^{inter}$ by a factor 2. The result obtained is displayed in Figure 5.

The full scheme is contained in Figure 3. This scheme in two steps is not essential to the method but it serves to speed up its convergence. We could also proceed to a direct minimization of (15) using equations (16), initialized with the image obtained from $\vec{u}^{3.2}$ by a simple replication of its pixels by a factor 4. Let us remark that more sophisticated initializations (FFT or DCT based, or based on bilinear interpolation) could be used and they lead to similar results than the ones reported here, though they exhibit a faster convergence when using a direct minimization, and a slightly faster convergence when using the two-step method.

The PSNR's of the R, G, B channels are $43.07, 46.55, 43.83$, respectively. Table 1 contains the statistics of errors between Figures 5 and 1. Note that the maximum error is 110.6 (recall that images take values in the range from 0 to 1023) in the red channel, and is localized in a pixel near the swimming pool, which has been lost due to the subsampling. High errors are mostly localized in regions with strong saturation of the intensity. Even if we do not display them here, let us mention that we also compared these results with the results obtained with a simple interpolation method like bilinear, and bicubique, and the maximum, ℓ^1, and ℓ^2 errors were typically 5 times bigger.

Finally, in Figure 6 we display the result obtained by minimizing (14) instead of (13). The PSNR of the R, G, B channels are $43.26, 46.74, 43.77$, respectively. The statistics of errors are displayed in Table 2. We see that the errors are slightly better. In spite of this, the Euler-Lagrange equations of functional (13) with $p = 2$ are linear in the unknown variables, and the gradient descent exhibits a faster convergence. The time spent reconstructing an image of size 800x800 pixels on a Pentium 1.8 GHz is of 63 seconds. A possible solution could be to compute first the result obtained using functional (13) with $p = 2$, followed by some gradient descent iterations of the equations corresponding to functional (14) in order to improve the result.

Figure 5: The reconstructed image obtained using functional (13) with $p = 2$.

Figure 6: The reconstructed image obtained using functional (14).

4 Conclusions

We have reported a variational model for increasing the resolution of satellite multispectral data knowing the panchromatic image at higher resolution and the multispectral data at a lower resolution. The model incorporates the relations between the spectral channels and the panchromatic image (1), and the relation describing how the low resolution pixel is formed from the high resolution pixel by a low pass filtering followed by subsampling (2). But the main feature of our model is the incorporation of the hypothesis that, for satellite multispectral images, *to a large extent, the geometry of the spectral channels is contained in the topographic map of its panchromatic image*. We have constructed two slightly different energy functionals (14) and (13) which incorporate the above three basic postulates. We have to note that functional (13) is invariant under contrast inversion, i.e, under the change of θ into $-\theta$. This could make it more adapted to treat the case of near infrared, red and green channels. Combination of both functional should also be explored. Finally, we described our algorithm to minimize them and we displayed some experiments.

Acknowledgement. This work was supported by the CNES.

	Errors								
	Max	L^1	L^2	Mode	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}
Red Channel	110,6	3,7	6,7	1	0,3	0,7	1,7	4,3	9,4
Trees/forest	8,8	0,9	1,2	1	0,1	0,3	0,7	1,3	1,9
Red roof houses	59,4	7,7	11,4	2	0,8	1,9	4,9	10,2	18,8
Gray roof houses	33,7	4,6	6,5	2	0,6	1,5	3,2	5,9	10,6
Swimming pools	110,6	21,9	29,3	2	1,5	6,7	18,5	30,4	46,2
Streets and ground	6,9	1,2	1,7	1	0,1	0,4	0,8	1,6	2,6
Red cars	47,1	24,4	28,3	#N/A	11,1	16,4	21,4	29,4	40,0
White dots of windows	81,0	30,9	37,2	38	3,9	10,7	31,6	44,4	64,1
Green Channel	101,8	2,4	3,9	1	0,2	0,6	1,4	3,0	5,6
Trees/forest	8,1	1,2	1,6	1	0,2	0,4	0,9	1,6	2,6
Red roof houses	36,8	4,2	6,2	1	0,5	1,1	2,6	5,4	9,8
Gray roof houses	15,2	1,9	2,8	1	0,3	0,6	1,3	2,5	4,2
Swimming pools	32,9	8,9	11,4	7	1,1	3,3	7,5	12,8	18,5
Streets and ground	4,2	0,7	1,0	1	0,2	0,3	0,5	1,0	1,5
Red cars	26,4	12,6	15,3	#N/A	4,3	7,6	10,9	15,9	22,2
White dots of windows	101,8	27,3	35,0	35	4,0	8,9	23,4	37,4	56,8
Blue Channel	77,8	2,5	4,3	1	0,2	0,6	1,4	3,1	6,1
Trees/forest	7,4	0,7	1,0	0	0,1	0,3	0,6	1,0	1,5
Red roof houses	30,3	4,3	6,1	1	0,5	1,3	3,0	6,0	10,0
Gray roof houses	18,5	3,4	4,6	1	0,6	1,3	2,6	4,5	7,2
Swimming pools	77,8	13,7	18,6	1	0,8	3,8	10,9	19,9	27,7
Streets and ground	4,7	1,0	1,2	1	0,2	0,4	0,8	1,3	1,9
Red cars	20,7	11,7	13,0	#N/A	6,7	8,7	10,4	13,4	17,8
White dots of windows	52,9	21,3	26,8	3	2,2	5,7	17,5	35,9	45,0

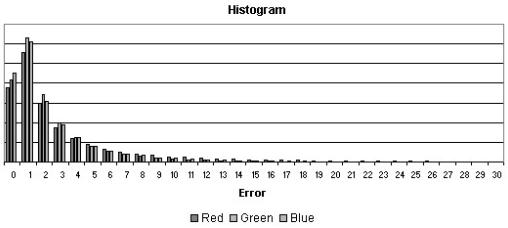

Table 1: errors corresponding to the experiment of Figure 5. In the histogram of errors, the columns correspond to red, green, and blue errors, respectively.

5 Appendix

Functions of bounded variation. Let Q be an open set in $\mathbb{R}^n$. A function $u \in L^1(Q)$ whose partial derivatives in the sense of distributions are measures with finite total variation in Q is called a function of bounded variation. The class of such functions will be denoted by $BV(Q)$. We say that a measurable set $E \subseteq Q$ has *finite perimeter* in Q if its indicator function $\chi_E \in BV(Q)$. If $u \in BV(Q)$ almost all its level sets $[u \geq \lambda] = \{x \in Q : u(x) \geq \lambda\}$ are sets of finite perimeter. For sets of finite perimeter E one can define the essential boundary $\partial^* E$, which is rectifiable with finite H^{n-1} measure, and compute the normal to the level set at H^{n-1} almost all points of $\partial^* E$. Thus at almost all points of almost all level sets of $u \in BV(Q)$ we may define a normal vector $\theta(x)$. This vector field of normals θ can be also defined (hence extended to all Q) as the Radon-Nikodym derivative of the measure ∇u with respect to $|\nabla u|$, i.e., it

	Errors								
	Max	L^1	L^2	Mode	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}
Red Channel	96.7	3.5	6.6	1	0.2	0.7	1.6	4.0	9.0
Trees/forest	8.4	0.9	1.3	1	0.1	0.3	0.7	1.3	1.9
Red roof houses	55.2	7.5	11.5	1	0.7	1.9	4.4	10.0	19.0
Gray roof houses	34.4	4.6	6.3	1	0.7	1.6	3.5	6.4	9.9
Swimming pools	86.3	17.9	24.2	1	1.5	5.2	14.2	25.9	40.4
Streets and ground	3.7	0.9	1.2	0	0.1	0.3	0.7	1.3	2.0
Red cars	39.0	18.9	22.6	16	8.3	13.0	15.7	21.6	32.0
White dots of windows	86.7	36.1	44.2	3	4.6	14.0	33.5	56.7	72.0
Green Channel	91.5	2.3	3.8	1	0.2	0.6	1.3	2.9	5.5
Trees/forest	10.5	1.2	1.7	1	0.2	0.4	0.8	1.6	2.5
Red roof houses	37.0	4.1	6.2	1	0.4	1.1	2.4	5.2	10.0
Gray roof houses	15.4	2.0	2.9	1	0.3	0.6	1.4	2.5	4.1
Swimming pools	37.9	8.6	11.4	1	1.0	2.8	6.7	12.5	19.0
Streets and ground	3.2	0.6	0.8	0	0.1	0.2	0.5	0.8	1.3
Red cars	22.3	10.2	12.6	8	3.7	6.7	8.4	11.9	18.1
White dots of windows	91.5	26.2	32.3	2	4.6	11.7	25.3	35.6	48.9
Blue Channel	83.3	2.5	4.4	1	0.2	0.5	1.3	3.0	6.0
Trees/forest	8.2	0.7	1.0	0	0.1	0.3	0.5	1.0	1.5
Red roof houses	33.1	4.4	6.4	1	0.5	1.2	2.9	5.9	9.7
Gray roof houses	19.1	3.4	4.4	2	0.5	1.3	2.7	4.6	7.1
Swimming pools	49.9	11.9	15.5	1	1.3	3.9	10.0	17.2	25.8
Streets and ground	3.5	0.9	1.1	1	0.2	0.3	0.7	1.1	1.8
Red cars	16.9	8.3	9.7	#N/A	4.0	4.9	6.4	9.8	14.0
White dots of windows	83.3	28.3	35.8	59	3.7	9.1	25.3	42.9	58.6

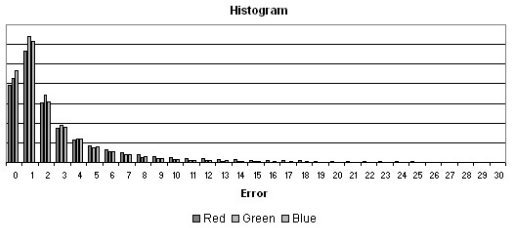

Table 2: errors corresponding to the experiment of Figure 6.

formally satisfies $\theta \cdot \nabla u = |\nabla u|$ and, also, $|\theta| \leq 1$ a.e.. We shall refer to the vector field θ as the vector field of unit normals to the topographic map of u. For further information concerning functions of bounded variation we refer to [3].

Existence of solutions of the variational problems (13) and (14). Let $W^{1,2}(\Omega)$ denote the space of functions $u \in L^2(\Omega)$ such that $\nabla u \in L^2(\Omega)$. Assume that $\theta : \Omega \to \mathbb{R}^2$ is such that $|\theta(x)| \leq 1$ a.e., and satisfies $div\, \theta^\perp \in L^2(\Omega)$. Let $W(\Omega, \theta)$ be the completion of $W^{1,2}(\Omega)$ with respect to the norm

$$\Phi(u) = \Big(\int_\Omega |\theta^\perp \cdot \nabla u|^2 \Big)^{1/2} + \Big(\int_\Omega |u|^2 \Big)^{1/2}.$$

We have the following result.

Theorem 1 *If $div\, \theta^\perp \in L^2(\Omega)$, then the functional (13) admits a minimum in $W(\Omega, \theta)^3$. Similarly, if $div\, \theta \in L^2(\Omega)$, functional (14) admits a minimum in $BV(\Omega)^3$.*

References

[1] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, Context-Driven Fusion of High Spatial and Spectral Resolution Images Based on Oversampled Multiresolution Analysis, IEEE Trans. on Geoscience and Remote Sensing, vol. 40, n. 10, pp. 2300–2312, 2002.

[2] L. Ambrosio, V. Caselles, S. Masnou and J.M. Morel, The Connected Components of Sets of Finite Perimeter, *European Journal of Math.*, 3, pp. 39-92, 2001.

[3] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, 2000.

[4] C. Ballester, V. Caselles, B. Rougé, and J. Verdera, *Une méthode geométrique de fusion des images P+XS*. Rapport CNES, 2003.

[5] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, *Filling-In by Joint Interpolation of Vector Fields and Gray Levels*, IEEE Transactions on Image Processing, Vol.10, n. 8, pp.1200-1211, 2001.

[6] V. Caselles, B. Coll and J.M. Morel, *Topographic Maps and Local Contrast Changes in Natural Images*, Int. J. Comp. Vision 33(1), pp. 5-27, 1999.

[7] V. Caselles, B. Coll and J.M. Morel, *Geometry and Color in Natural Images*, J. Math. Imaging and Vision, 16, pp. 89-107, 2002

[8] F. Guichard and J.M. Morel, *Partial differential equations and image iterative filtering*, Forthcoming book.

[9] J.L. Lisani and J.M. Morel, *Comparaison morphologique d'images*, Rapport CNES, 2000.

[10] L.I. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms.* Physica D, **60**, 259–269.

[11] J. Serra, *Image analysis and mathematical morphology.* Academic Press, 1982.

[12] L. Wald, T. Ranchin, and M. Mangolini, Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images, Photogramm. Eng. Remote. Sensing, vol. 63(6), pp. 691–699, 1997.

[13] M. Wertheimer, Untersuchungen zur Lehre der Gestalt, II, *Psychologische Forschung*, 4, 1923, pp. 301–350.