

# Constrained Anisotropic Diffusion and some Applications

Gabriele Facciolo<sup>1,3</sup>, Federico Lecumberry<sup>1</sup>, Andrés Almansa<sup>1</sup>,  
Alvaro Pardo<sup>1,2</sup>, Vicent Caselles<sup>3</sup>, Bernard Rougé<sup>4</sup>

1 - IIE, INCO, Facultad de Ingeniería, UdelaR, Montevideo, Uruguay

2 - DIE, Universidad Católica del Uruguay, Montevideo, Uruguay

3 - Departament de Tecnologia, Universitat Pompeu Fabra, Barcelona, Spain

4 - CNES, Centre National d'Etudes spatiales, Toulouse, France

{facciolo, fefo, apardo, almansa}@fing.edu.uy,

vicent.caselles@upf.edu, bernard.rouge@cnes.fr

## Abstract

Minimal surface regularization has been used in several applications ranging from stereo to image segmentation, sometimes hidden as a graph-cut discrete formulation, or as a strictly convex approximation to TV minimization. In this paper we consider a modified version of minimal surface regularization coupled with a robust data fitting term for interpolation purposes, where the corresponding evolution equation is constrained to diffuse only along the isophotes of a given image  $u$  and we design a convergent numerical scheme to accomplish this. To illustrate the usefulness of our approach, we apply this framework to the digital elevation model interpolation and to constrained vector probability diffusion.

## 1 Introduction

In our work we apply a PDE based anisotropic diffusion technique that imposes both geometric and data fit constraints to interpolation of digital elevation models (DEM). Interpolation of DEM models requires to diffuse the known data points to unknown areas, mainly because the elevation information is not valid everywhere. There are plenty of techniques to obtain DEMs but we assume that the computation of the model has been already done but the result is sparse and/or noisy. We propose to interpolate the unknown areas by constraining a diffusion process to the geometry imposed by a reference image, and coupling the process with a data fitting term which tries to adjust the reconstructed surface to the known data. In a similar way this technique could be applied to the diffusion probability (as we will show later), or to other denoising and restoration problems.

For simplicity, we shall work in a discrete framework, and thus images are represented as vectors in  $X := \mathbb{R}^{N \times N}$ . The above problem can be modelled as an optimization problem:

$$\min_{p \in X} (E(p) + \lambda D(p)) \quad (1)$$

where  $p \in X$ ,  $D(p)$  is a term of attachment to the data that will be described later, and

$E(p)$  is an anisotropic regularization energy written as:

$$E(p) = \sum_{i,j=1}^N \sqrt{\beta^2 + |A_{(i,j)} \nabla p_{(i,j)}|^2} \quad (2)$$

where  $A_{(i,j)}$  is a  $2 \times 2$  matrix for each  $(i, j)$  and  $\nabla$  denotes any discretization of the gradient (to be made precise later). The matrix  $A$  is called the *diffusion tensor* and allows to embed some information on the directions we want to constrain the diffusion process. The proposed equation (2) belongs to a family that has proven to be very useful in surface regularization [8]; a representative of such family, even if less suited in the present case, corresponds to  $A = Id$ , in this case the continuous analogue of the energy (2) measures the surface area of  $p$  (where  $\beta$  is the ratio between the horizontal dimensions and the vertical dimension).

The gradient descent strategy to minimize (1) leads to the equation:

$$\frac{\partial p}{\partial t} = \text{div}^{--} \left( \frac{A^\dagger A \nabla^{++} p}{\sqrt{\beta^2 + |A \nabla^{++} p|^2}} \right) - \lambda \frac{\partial D}{\partial p}(p), \quad (3)$$

(where  $-\text{div}$  denotes the dual of the gradient  $\nabla$  in  $X$ ) plus Neumann boundary conditions. Diffusion equations of the form:

$$\partial_t p = \text{div}(\mathcal{A} \nabla p) \quad (4)$$

where  $\mathcal{A}$  is a diffusion tensor, i.e., a map which associates to each  $(i, j)$  a positive definite matrix  $\mathcal{A}_{(i,j)}$  have been studied in many contexts. In particular, Weickert [14] studied the continuous, the time discrete and spatio-temporal discrete tensor diffusion models and its applications; Tschumperle considered this problem in [13] together with its applications to the diffusion of multi-valued data. In [15] Welk et al. have developed the so called Locally Analytic Schemes (LAS) for this type of problems and analyzed its connections to wavelet shrinkage.

Even if the general model (4) contains (3) with  $\lambda = 0$  as a particular case, the structure tensor that we use does not satisfy the restrictions imposed in [14] to develop a stable numerical approach (some diffusion across the level-lines is required in order to obtain a stable numerical approximation). On the other hand, the precise form of  $\mathcal{A}$  that we are going to use is adapted to the applications we have in mind.

We will develop a globally convergent and explicit numerical scheme for the solution of (3). The use of implicit methods is not always convenient; this may be the case when the data term is too complex, like the *fundamental equation of correlation* (proposed in [6]) for which only an approximation is available [7]. Other solutions based on energy minimization via graph cuts could be used [3] but they may require an excessive memory due to the size of our images (a digital elevation model may consist of images of size  $1000 \times 1000$  pixels with a fine quantization of possible altitude values).

Let us describe the plan of the paper. In Section 2 we describe the precise model used for DEM interpolation: we make explicit the structure tensor and the data fitting terms. In Section 3 we propose (and prove the convergence of) an explicit numerical scheme for the solution of (3) when the data fitting term is of robust type. In Section 4 we apply the proposed method to the interpolation of digital elevation models and we display some experiments to illustrate its performance. Section 5 is devoted to study the anisotropic vector probability diffusion model (VPD). Finally, Section 6 summarizes the main conclusions of the paper.

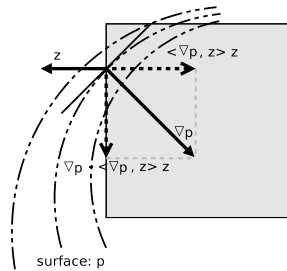


Figure 1: Regularization subject to level lines. The evolution of surface  $p$  (dotted level set), is subject to the geometry (square) whose normals are  $z$ . The gradient of  $p$  is projected over the basis  $\{z, z^\perp\}$  and  $\nabla p - \langle \nabla p, z \rangle z$  is aligned to  $z^\perp$ .

## 2 An anisotropic diffusion model for DEM regularization and interpolation

The windowed correlation is at the base of the most precise [6] photogrammetric techniques but generates elevation maps that are sparse and irregular (Fig. 2). It seems necessary to impose some regularity, specially in the reconstruction of urban areas where the height map is essentially given by a piecewise affine model and where the relevant information is mostly concentrated on the borders of the objects.

Assume that  $u \in X$  is a given reference image of a stereo pair from which we have computed the disparity map and the associated height reconstruction  $m: \{1, \dots, N\}^2 \rightarrow \mathbb{R}$ . We also dispose of a mask function  $w_{(i,j)}$  which is 1 at the points  $(i,j)$  for which we dispose of an acceptable estimate of the error standard deviation ( $Err_{(i,j)}$ ) and 0 otherwise (see [6]). It may happen that the known data  $m$  is sparse, specially after considering the mask  $w$ . In that case we want to interpolate them to get a reconstructed height function  $p$  on the full grid by minimizing a energy functional of the form (1). Assuming that the geometry of the reference image  $u$  (expressed in terms of its edges, representing buildings in case of urban area images) is correlated to the height  $p$ , we shall use an structure tensor that correlates both the geometries of  $u$  and  $p$ . For that, as in [4, 13], we diffuse the image  $p$  along the level lines of  $u$  that serves as a geometric constraint. To do this, we use the vector field  $z$  of unit normals to the level lines of  $u$  and we inhibit the diffusion of  $p$  across those level lines, in other words, we inhibit diffusion of  $p$  in the direction of  $z$ . For that we define the structure tensor  $A_{(i,j)} = (I - z_{(i,j)} \otimes z_{(i,j)})$  so that

$$A_{(i,j)}(\nabla p_{(i,j)}) = \begin{pmatrix} 1 - z_x^2 & -z_x z_y \\ -z_x z_y & 1 - z_y^2 \end{pmatrix} \nabla p_{(i,j)} = \nabla p_{(i,j)} - \langle z_{(i,j)}, \nabla p_{(i,j)} \rangle z_{(i,j)} \quad (5)$$

where the entries of the matrix are evaluated at  $(i,j)$ . We define a regularized version of the vector field of unit normals to the level lines of  $u$  in the following way: we define  $z_{(i,j)} = \frac{\nabla u_{(i,j)}}{|\nabla u_{(i,j)}|}$  when  $|\nabla u_{(i,j)}| \geq b$  and  $z_{(i,j)} = \frac{\nabla u_{(i,j)}}{\sqrt{b^2 + |\nabla u_{(i,j)}|^2}}$  otherwise, where  $b > 0$  is a threshold on the gradient which ensures that its direction is reliable, in our cases we select  $b = 50$ , for images with a range  $[0, 3000]$ .

Concerning the data fitting term, we have explored several possibilities. The most common is the weighted  $L^2$  norm  $D_2(p, m) = \sum_{i,j=1}^N w_{(i,j)} d_2(p_{(i,j)}, m_{(i,j)})$ , where  $w_{(i,j)}$

is the binary function described previously and  $d_2(x, y) = (x - y)^2$ . When the data  $m$  are contaminated by outliers, it is more useful to use a robust term like the  $L^1$  norm, or the following variant:  $D_1(p, m) = \sum_{i,j}^N w_{(i,j)} d_1(p_{(i,j)}, m_{(i,j)})$ ,  $d_1(x, y) = \sqrt{a^2 + (x - y)^2}$ ,  $a > 0$ , which is chosen here because it keeps the robustness of the  $L^1$  norm and avoids the numerical difficulties associated to the non-differentiability of the  $L^1$  norm at 0. Generally we may dispose of an estimate of the noise standard deviation ( $\sigma$ ) that affects the measures of  $m$ , then we may impose the data fitting term as a constraint  $D_2(p, m) \leq \sigma^2 L$  where  $L$  is the number of pixels at which  $w_{(i,j)} \neq 0$  (or a similar one for  $D_1(p, m) \leq \sqrt{a^2 + \sigma^2 L}$ ) and choose the value of  $\lambda$  so that the constraint is satisfied [2, 9, 1] (for instance, by means of Uzawa's algorithm).

But using a single Lagrange multiplier will give an unfair weight to the values corresponding to "higher ground" in the computation of the error. In the case of DEM reconstruction, since the disparities were calculated from a stereo pair by correlation, we may assume that the differences  $(p - m)$  have 0 mean, and that we have at our disposal a local estimate  $Err(m)_{(i,j)}^2$  of the noise variance at each point  $(i, j)$ . As it has been shown in [6], such estimate (named *correlation curvature*) is related to the correlation process and suggests that the measures obtained near the image borders are the most reliable. Including this information directly into the data fitting terms we obtain two new data terms:

$$D_2^{**}(p, m) = \sum_{i,j=1}^N w_{(i,j)} d_2^{**}(p_{(i,j)}, m_{(i,j)}, Err_{(i,j)}) \quad \text{where} \quad d_2^{**}(x, y, e) = \frac{(x - y)^2}{e^2} \quad (6)$$

$$D_1^{**}(p, m) = \sum_{i,j=1}^N w_{(i,j)} d_1^{**}(p_{(i,j)}, m_{(i,j)}, Err_{(i,j)}) \quad \text{where} \quad d_1^{**}(x, y, e) = \sqrt{1 + \frac{(x - y)^2}{e^2}} \quad (7)$$

In both cases the new expected error per pixel is 1, and we write the constraints as:  $D_2^{**}(p, m) \leq L$  and  $D_1^{**}(p, m) \leq \sqrt{2}L$ . This observation allows us to use a unique value of the multiplier  $\lambda$  for the whole data term; the resulting formulation is closely related to the one presented in [9] but these authors adapt locally the values of their parameters  $\lambda$ .

### 3 Numerical Analysis of Constrained Diffusion

We denote by  $X$  the Euclidean space  $\mathbb{R}^{N \times N}$ . The Euclidean scalar product and the norm in  $X$  will be denoted by  $\langle \cdot, \cdot \rangle_X$  and  $\| \cdot \|_X$ , respectively. Then the image  $u \in X$  is the vector  $u = (u_{(i,j)})_{i,j=1}^N$ . If  $u \in X$ , the discrete gradient is a vector in  $Y = X \times X$  given by  $\nabla^{++} u := (\nabla_x^+ u, \nabla_y^+ u)$ , where for each  $i, j = 1, \dots, N$  we define  $\nabla_x^+ u_{(i,j)} = u_{(i+1,j)} - u_{(i,j)}$  if  $1 \leq i < N$  and  $\nabla_x^+ u_{(i,j)} = 0$  if  $i = N$ , with an analogous expression for  $\nabla_y^+ u_{(i,j)}$ . Other choices of the gradient are possible, this one will be convenient for the developments below. We denote the euclidean norm of a vector  $v \in \mathbb{R}^2$  by  $|v|$ .

The Euclidean scalar product in  $Y$  is defined in the standard way by  $\langle \vec{p}, \vec{q} \rangle_Y = \sum_{i,j=1}^N (p_{i,j}^1 q_{i,j}^1 + p_{i,j}^2 q_{i,j}^2)$  for every  $\vec{p} = (p^1, p^2)$ ,  $\vec{q} = (q^1, q^2) \in Y$ . The norm of  $\vec{p} = (p^1, p^2) \in Y$  is, as usual,  $\|\vec{p}\|_Y = \langle \vec{p}, \vec{p} \rangle_Y^{1/2}$ . By analogy with the continuous setting, we introduce a discrete divergence  $\text{div}^{--}$  as the dual operator of  $\nabla^{++}$ , i.e., for every  $\vec{p} \in Y$  and  $u \in X$  we have

$$\langle -\text{div}^{--} \vec{p}, u \rangle_X = \langle \vec{p}, \nabla^{++} u \rangle_Y.$$

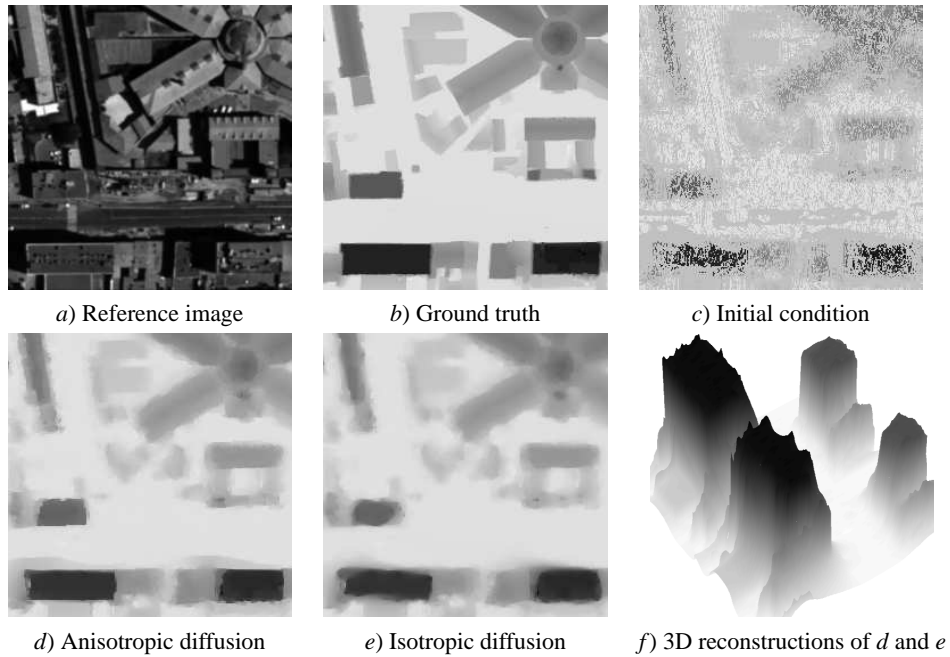


Figure 2: DEM interpolation. From left to right: a sparse DEM to be interpolated, the ground truth information, the reference image, two results obtained after regularization with anisotropic and isotropic terms (using  $D_1^{**}$  data fitting) and the reconstructions side by side of them (front row is the isotropic, while the back row is anisotropic). The elevation is encoded as a gray level images.

One can easily check that  $\text{div}^{--}$  is a discrete version of the divergence computed using backward derivatives (see [5]).

Let us consider the discrete functional:

$$E(p) + \lambda D(p) = \sum_{1 \leq i, j \leq N} \sqrt{\beta^2 + |A_{(i,j)} \nabla^{++} p_{(i,j)}|^2} + \lambda \sum_{1 \leq i, j \leq N} w_{(i,j)} \alpha_{(i,j)} \sqrt{a^2 + |p_{(i,j)} - m_{(i,j)}|^2} \quad (8)$$

where  $w_{(i,j)}$  is a binary mask defined in Section 2 and  $\alpha_{(i,j)} > 0$  is a weighting function to be described later. We propose to solve the optimization problem:

$$\min_{p \in X} E(p) + \lambda D(p). \quad (9)$$

Since the functional  $E + \lambda D$  is strictly convex, there is a unique minimum of it in  $X$ , call it  $p^*$ . To minimize  $E(p) + \lambda D(p)$  we use an explicit gradient descent method, i.e., an iterative scheme of the form  $p^{n+1} = p^n - t_n \nabla(E + \lambda D)(p^n)$  with  $t_n > 0$  for all  $n$ , which in the present case is:

$$p^{n+1} = p^n + t_n \left( \text{div}^{--} (A_{p^n}^t A_{p^n} \nabla^{++} p^n) - \lambda w \alpha \frac{p^n - m}{\sqrt{a^2 + |p^n - m|^2}} \right), \quad (10)$$

plus Neumann boundary conditions, where  $A_p$  is defined as  $A_p = A_p^t = \frac{A}{(\beta^2 + |\nabla^{++} p|^2)^{\frac{1}{4}}}$ .

Observe that, since we are in the discrete case, the term inside the parenthesis is bounded in  $X$  with a bound independent of  $n$ . Moreover, since the minimum of  $E + \lambda D$  is unique, then using Theorem 1 below (see [12])  $\{p^n\}$  converges to  $p^*$

**Theorem 1** *Let  $X$  be a vector space of finite dimension and  $J$  be a convex function defined on  $X$  which has a bounded set of minimum points  $X^*$ . Assume that  $t_n$  is a sequence of positive numbers satisfying the conditions*

$$t_n > 0, \quad \lim_{n \rightarrow +\infty} t_n = 0, \quad \sum_{n=0}^{+\infty} t_n = +\infty, \quad (11)$$

and let  $x^k$  be a sequence of vectors in  $X$  generated by the formula:  $x^{k+1} = x^k - t_k \partial J(x^k)$ , where  $\partial J(x_k)$  is a subgradient of  $J$  at  $x^k$  with initial condition given by some  $x^0 \in X$ . If the sequence  $\partial J(x^k)$  is bounded, then there is a subsequence converging to some point  $x^* \in X^*$ . If the minimum of  $J$  is unique, then  $x^k$  converges to it.

The above formulation (9) covers several interesting case:

- **Case 1:**  $\alpha = 0$ . In this case we are only considering the energy  $E(p)$ , and its evolution is given by:  $p^{n+1} = p^n + t_n \operatorname{div}^{--} (A_{p^n}^t A_{p^n} \nabla^{++} p^n)$ .
- **Case 2:**  $\alpha = 1$ ,  $a$  being a positive constant and  $w \in X$  being a binary mask. It corresponds to the minimization of  $E(p) + \lambda D_1(p)$  described in Section 2.
- **Case 3:**  $\alpha = \frac{1}{\operatorname{Err}(m)}$ ;  $a = \operatorname{Err}(m)$ . In this case the value of  $\alpha = 1/a$  is not constant and can be interpreted as a weighting function to impose a local control on the errors. In this case, we minimize the functional  $E(p) + \lambda D_1^{**}(p)$ .

**Remark 1** Now let us consider the case where the data term in the energy is quadratic, for instance, assume that we replace the term  $D(p)$  with:

$$D_2(p) = \sum_{1 \leq i, j \leq N} \alpha_{(i,j)} w_{(i,j)} (p_{(i,j)} - m_{(i,j)})^2 \quad (12)$$

and we minimize  $E(p) + \lambda D_2(p)$ . In this case,  $\frac{\partial D_2}{\partial p} = 2\alpha_{(i,j)} w_{(i,j)} (p^n - m)$  is not bounded  $\forall p^n$  and we cannot use Theorem 1. Let us denote

$$\mathcal{C}_n(p) := \operatorname{div}^{--} (A_{p^n}^t A_{p^n} \nabla^{++} p), \quad B_n(p) = \mathcal{C}_n(p) - 2\lambda \alpha w p$$

where  $\alpha w p_{(i,j)} = \alpha_{(i,j)} w_{(i,j)} p_{(i,j)} \forall (i, j)$ . In this case, we may write the gradient descent equations for  $E(p) + \lambda D_2(p)$  as:

$$p^{n+1} = p^n + dt \mathcal{C}_n(p^n) - dt \lambda \alpha w 2(p^n - m) = (Id + dt B_n) p^n + 2dt \lambda \alpha w m. \quad (13)$$

By iterating we may express  $p^{n+1}$  in terms of  $p^0$  by the formula:

$$p^{n+1} = \underbrace{\prod_{i=0}^n [Id + dt B_i(p^i)]}_{I} p^0 + 2dt \lambda \alpha w m \underbrace{\left\{ 1 + \sum_{j=0}^{n-1} \left( \prod_{k=j+1}^n [Id + dt B_k(p^k)] \right) \right\}}_{II} p^0.$$

The sequence  $p^n$  is bounded and the above method is stable. This will follow as a consequence of Proposition 1 below which guarantees the boundedness of the term (I) and the convergence of the series in (II) as soon as  $dt < \frac{\beta}{4+\lambda\alpha\beta}$ . This condition allows to cover two cases:

- **Case A:**  $\alpha = 1$  and  $w \in X$  being a binary mask. It corresponds to the minimization of  $E(p) + \lambda D_2(p)$  described in Section 2.
- **Case B:**  $\alpha = \frac{1}{Err(m)^2}$ . In this case the value of  $\alpha$  is not constant and can be interpreted as a weighting function to impose a local control on the errors. In this case, we minimize the functional  $E(p) + \lambda D_2^{**}(p)$ .

For a linear operator  $A : X \rightarrow X$ , we denote by  $\rho(A)$  its spectral radius, that is  $\max\{|\lambda_i| : \lambda_i \text{ eigenvalue of } A\}$ .

**Proposition 1** *If  $dt < \frac{\beta}{4+\lambda\alpha\beta}$ , then  $\|Id + dtB_n\|_X = \rho(Id + dtB_n) \leq 1 - \varepsilon$  for some  $\varepsilon > 0$ .*

**Proof.** Observe that  $\mathcal{C}_n$  is a symmetric operator and thus its norm in  $X$  coincides with its spectral radius. Now, observe that  $\lambda_i$  is an eigenvalue of  $Id + dtB_n$  if and only if  $\lambda_i = 1 + dt\mu_i - 2dt\lambda\alpha$  for some eigenvalue  $\mu_i$  of  $\mathcal{C}_n$ . Our statement follows if we prove that

$$-1 + \varepsilon \leq 1 + dt\mu - 2dt\lambda\alpha \leq 1 - \varepsilon \quad \forall \mu \text{ eigenvalue of } \mathcal{C}_n. \quad (14)$$

Now, we observe that the eigenvalues of  $\mathcal{C}_n$  are contained in the interval  $[-\frac{8}{\beta}, 0]$ . Indeed, if  $x$  is an eigenvector of  $\mathcal{C}_n$  corresponding to the eigenvalue  $\mu$ , we have

$$\begin{aligned} \mu \|x\|_2^2 = \langle \mathcal{C}_n(x), x \rangle_X &= \langle \text{div}^{--}(A_x^t A_x \nabla^{++} x), x \rangle_X = -\langle A_x^t A_x \nabla^{++} x, \nabla^{++} x \rangle_Y \\ &= -\langle A_x \nabla^{++} x, A_x \nabla^{++} x \rangle_Y = -\|A_x \nabla^{++} x\|_Y^2 \leq 0 \end{aligned}$$

and  $\|A_x \nabla^{++} x\|_Y^2 \leq \frac{1}{\beta} \|A \nabla^{++} x\|_Y^2 \leq \frac{8}{\beta}$  because  $\|A_x\| \leq 1$  and  $\|\nabla^{++} p\|_Y^2 \leq 8\|p\|_X^2$ . Since the eigenvalues of  $\mathcal{C}_n$  are negative, (14) can be written as

$$2 - \varepsilon \geq dt(|\mu| + 2\lambda\alpha) \geq \varepsilon. \quad (15)$$

If  $\{\mu_i\}$  are the eigenvalues of  $\mathcal{C}_n$ , this condition is implied if we choose  $dt$  such that

$$\frac{\varepsilon}{\min_i |\mu_i| + 2\lambda\alpha} \leq dt \leq \frac{1 - \varepsilon}{\max_i |\mu_i| + 2\lambda\alpha},$$

which is implied by a choice of  $dt$  such that  $\frac{\varepsilon}{2\lambda\alpha} \leq dt \leq \frac{1 - \varepsilon}{(8/\beta) + 2\lambda\alpha}$ . We can find a value of  $\varepsilon > 0$  satisfying this as soon as  $dt < \frac{\beta}{4+\lambda\alpha\beta}$ .

## 4 Experimental results

The experiments with DEM interpolation were performed using a low disparity synthetic data set kindly provided by CNES - Centre National d'Etudes spatiales (the reference image is shown in figure 2a). It was also provided a disparity map (figure 2c) and the map of estimated disparity errors for each point ( $Err_{(i,j)}$ ), both were obtained according to

DATA TERMS	ISOTROPIC		ANISOTROPIC	
	Mean $d_1$	Mean $d_2$	Mean $d_1$	Mean $d_2$
$D_1$	1.0165	0.0367	1.0106	0.0236
$D_1^{**}$	1.0143	0.0321	1.0115	0.0254
$D_2$	1.0199	0.0438	1.0118	0.0259
$D_2^{**}$	1.0158	0.0358	1.0123	0.0274
<i>initial condition</i>	1.0241	0.0552	N/A	N/A

Table 1: Experiments with DEM interpolation, here are reported different combination of interpolation algorithm and data terms. The errors are measured against the ground truth information as the mean of  $d_1$  and the mean of  $d_2$ .

the correlation technique described in section 2. As can be seen in figure 2c the initial disparity map is composed by sparse information, mostly because some points have errors beyond a certain threshold, so the known data must be interpolated. To test the performance of the different interpolation strategies we compare them against the provided ground truth (figure 2b).

All the experiments where run with the same altitude ratio  $\beta = 0.025$ , the time step  $dt$  was determined according to the described in Section 3, and the  $\lambda$  parameter is adjusted to using the Uzawa's algorithm [2, 1] to satisfy the constraints corresponding to each data term (section 2). The local and the global constraints can be related by the following observation, knowing the error map  $Err_{(i,j)}$  (which is the standard deviation of each point) the global standard deviation is determined as  $\sigma = \sqrt{\frac{1}{L} \sum_{i,j=0} Err_{(i,j)}^2} = 0.032$ .

In figure 2 we compare the results of isotropic (figure 2e) and anisotropic (figure 2d) regularization of the data given in figure 2c. As it can be seen, the shape driven regularization produces a more coherent result. This is confirmed by the reconstruction displayed in figure 2f and by the errors in table 1 (computed with respect to the ground truth). We have also verified that the data fitting constraints are satisfied by our algorithm and that the data fitting terms have a better local behavior when using  $D_i^{**}$  that when using  $D_i$ . Finally let us observe that the use of a robust norm ( $L^1$ ) effectively increases the precision, but this improvement is generally marginal compared to the obtained with the use of the precision information ( $Err(m)$ ).

## 5 A Model for Anisotropic Vector Probability Diffusion

The diffusion of probability vector maps (VPD), introduced in [11] as a tool for improving classification based on the same probability vector, is based on a vectorial form of the Total Variation functional. An anisotropic version of it was proposed in [10] in order to stop the diffusion across the edges of an image  $u$  while allowing the diffusion along them. This is achieved by means of the tensor  $A$  defined in (5), then given a vector of probabilities  $\vec{p}_{(i,j)} \in \mathcal{P} = \{\vec{p} \in \mathbb{R}^n : \|\vec{p}\|_1 = 1, p_k \geq 0\}$  at each pixel  $(i, j)$ , the MVPD formulation results in:

$$\frac{\partial p_k}{\partial t} = \text{div} \left( \frac{A^\dagger A \nabla p_k}{\sqrt{\beta^2 + \sum_{i=1}^n \|A \nabla p_k\|^2}} \right) \quad k = 1, \dots, n. \quad (16)$$



The original formulation of VPD corresponds to the choices  $A = Id$  and  $\beta = 0$  in (16).

Observe that, to be consistent, the solution of (16) should be in the simplex  $\mathcal{S}$ , i.e.,  $\vec{p}^t(i,j) \in \mathcal{S}$  for each  $(i,j)$  and any  $t > 0$ . At the continuous level, this would be guaranteed by the equation (16). But at the discrete level, if we use an explicit numerical scheme the constraints are not always satisfied (for instance, it is not difficult to check that the maximum principle is violated), and have to be imposed after each iteration. For that we define the projection  $\mathbb{P}$  of any vector  $q \in \mathbb{R}^n$  onto the simplex  $\mathcal{S}$  as the unique solution  $\mathbb{P}(q)$  of the minimum distance problem  $\min_{p \in \mathcal{S}} \|q - p\|^2$ . Then we consider the following discrete approximation to (16):

$$p^{\vec{n}+1} = \mathbb{P} \left[ \vec{p}^n + dt \operatorname{div} \left( \frac{A^\dagger A \nabla \vec{p}^n}{\sqrt{\beta^2 + \sum_{i=1}^n \|A \nabla \vec{p}^n\|^2}} \right) \right] \quad (17)$$

It is easy to prove that the equation (17) satisfies the conditions of theorem 1, proving that the associated method is stable and convergent. The method falls in the first analyzed case of section 3 where the data term of (1) is disabled by taking  $\lambda = 0$ .

An example of the application to the classification problem is shown in figure 3, there are compared the classification errors of VPD and MVPD against the manual segmentation of the same frame, as well as the probability level lines [10].

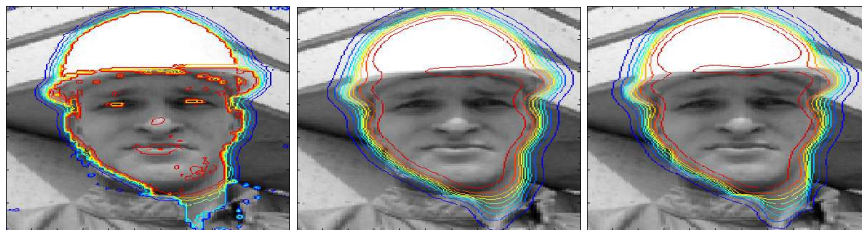


Figure 3: Results of probability diffusion, plotted as level lines over the image. At left is plotted the initial probability map which has an error of 82 miss-classified pixels (of the 6281 pixels belonging to the manually segmented region), centered is the result of VPD (49 erroneous pixels) and at right MVPD (40 erroneous pixels). It can be noticed that the level lines separating the helmet from the face in MVPD image tend to follow the borders while the VPD does not.

## 6 Conclusions and Future Work

We have considered a variational model which couples an anisotropic regularization term with a data fitting term and its applications to the interpolation of sparse digital elevation models and to probability diffusion. To minimize the energy functional, we solve its Euler-Lagrange equation via a gradient descent approach using an explicit numerical scheme, and we prove its convergence to the global minimum in the case of the  $L^1$  based data terms and its stability in the case of the  $L^2$  data terms. We have observed that the use of an anisotropic diffusion (respecting the geometry of a reference image) and the use of a robust data fitting term permits to improve the results when interpolating urban digital elevation models.

Several tasks remain for the future: the design of a fast numerical scheme, the evaluation of other techniques like Graph-Cuts, and the design of a multiresolution numerical scheme. The incorporation of different data terms like the proposed in [6] is expected to improve the results and further study of the local data fitting term model.

## References

- [1] A. Almansa, C. Ballester, V. Caselles, and G. Haro. A TV based restoration model with local constraints. *Preprint*, 2006.
- [2] M. Bertalmío, V. Caselles, B. Rougé, and A. Solé. TV based image restoration with local constraints. *Journal of Scientific Computing*, 19(1-3):95–122, 2003.
- [3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23(11):1222–1239, 2001.
- [4] B.Rougé C.Ballester, V.Caselles and J.Verdera. A variational model for p+xs image fusion. *Int. J. Comp. Vision*, 69:143–158, 2006.
- [5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20(1-2):89–97, 2004.
- [6] J. Delon. *Fine comparison of images and other problems*. PhD thesis, École Normale Supérieure de Cachan, 2004.
- [7] G. Facciolo, A. Almansa, and A. Pardo. Variational approach to interpolate and correct biases in stereo correlation. In *Proc. GRETSI Conference*, volume 2, pages 513–516, Louvain-La-Neuve, Belgium, sept 2005.
- [8] O. Faugeras and R. Keriven. Variational principles, surface evolution, pde's, level set methods, and the stereo problem. *T-IP*, 7:336–344, 1998.
- [9] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Variational denoising of partly-textured images by spatially varying constraints. *accepted for publication, IEEE TIP*, 2005.
- [10] F. Lecumberry. Cálculo de disparidad y segmentación de objetos en secuencias de video. Master's thesis, Facultad de Ingeniería, UdelaR, 2005.
- [11] Alvaro Pardo and Guillermo Sapiro. Vector Probability Diffusion. *IEEE Signal Processing Letters*, 8(4):106–109, April 2001.
- [12] N.Z. Sohr. *Minimization Methods for Non-differentiable Functions*. Springer, 1985.
- [13] D. Tschumperlé. *PDE-Based Regularization of Multivalued Images and Applications*. PhD thesis, University of Nice-Sophia Antipolis, December 2002.
- [14] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart, Germany, 1998.
- [15] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Technical Report No. 2100, IMA, University of Minnesota, Minneapolis, USA, February 2006.