

# Solving Ill-posed Problems Using Data Assimilation. Application to optical flow estimation

Dominique Béréziat, Isabelle Herlin

► **To cite this version:**

Dominique Béréziat, Isabelle Herlin. Solving Ill-posed Problems Using Data Assimilation. Application to optical flow estimation. VISAPP - International Conference on Computer Vision Theory and Applications, Feb 2009, Lisboa, Portugal. pp.594-602. inria-00567444

**HAL Id: inria-00567444**

**<https://hal.inria.fr/inria-00567444>**

Submitted on 8 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SOLVING ILL-POSED PROBLEMS USING DATA ASSIMILATION. APPLICATION TO OPTICAL FLOW ESTIMATION

Dominique Béréziat

*Université Pierre et Marie Curie, LIP6, 4 place Jussieu, 75005 Paris, France*

*Dominique.Bereziate@upmc.fr*

Isabelle Herlin

*INRIA, CERA, Joint Laboratory ENPC-EDF R&D, Université Paris-Est*

*BP105 78153 Le Chesnay Cedex, France*

*Isabelle.Herlin@inria.fr*

**Keywords:** Data Assimilation, Ill-posed Problem, Variational Formulation, Optical Flow, Missing Data.

**Abstract:** Data Assimilation is a mathematical framework used in environmental sciences to improve forecasts performed by meteorological, oceanographic or air quality simulation models. Data Assimilation techniques require the resolution of a system with three components: one describing the temporal evolution of a state vector, one coupling the observations and the state vector, and one defining the initial condition. In this article, we use this framework to study a class of ill-posed Image Processing problems, usually solved by spatial and temporal regularization techniques. A generic approach is defined to convert an ill-posed Image Processing problem in terms of a Data Assimilation system. This method is illustrated on the determination of optical flow from a sequence of images. The resulting software has two advantages: a quality criterion on input data is used for weighting their contribution in the computation of the solution and a dynamic model is proposed to ensure a significant temporal regularity on the solution.

## 1 Introduction

In the research field of Image Processing, most problems are ill-posed in the sense that it is not possible to provide a unique solution (Hadamard, 1923). A common cause of ill-posedness is that the equations used to model image properties, called Image Model, are under-determined. An example is the famous “aperture problem” occurring in the estimation of optical flow.

A possible strategy to solve ill-posed problems is to provide the Image Model with additional information. Two options are considered: 1) Providing explicit information: additional images are used to enlarge the set of input data. However, this is generally not possible because other acquisitions having the requested properties are often not available. 2) Providing implicit information: either hypotheses on image properties or constraints on the solution can be used. A usual constraint is to restrict the dimension or the size of the space of admissible solutions: for instance the result is searched among the functions having bounded spatial variations (Tikhonov, 1963). Such methods are called “Tikhonov regularization methods”. In the general case, these additional image

properties and constraints are expressed as equations, which, added to the Image Model, lead to a new invertible model.

Assuming an image sequence is available, the Tikhonov regularization can be extended by considering a regularization acting both in spatial and temporal domains. This has been proposed by Weickert (Weickert and Schnörr, 2001) for determining optical flow. The interest is to produce a solution which is regular in time and in space. This approach has however several drawbacks. First, the temporal regularization imposes an arbitrary regularity which can not deal with complex dynamics. Second, if irrelevant input data values occur on large regions, they are however taken into account in the estimation process and introduce errors in the solution. These aberrant values may be due to a sensor failure or an acquisition noise. We call them “missing data” in the remain of this paper. Third, as the solution is looked for in the spatio-temporal domain, this leads to a high complexity in comparison with a pure spatial model, which is a limiting factor for operational purposes.

An alternative to the Weickert’s approach is to model the temporal evolution of images. Such a modelling can be inferred, for instance, from *a priori*

knowledge on image acquisition. The challenge becomes to define an effective and pertinent model, and to include it in the solution computation. Data quality has also to be evaluated in order to ignore missing data.

In this paper, we propose to use the *Data Assimilation* framework as a generic tool to solve ill-posed problems in Image Processing. A Data Assimilation method solves simultaneously an evolution equation, describing the evolution of a state vector over time, and an observation equation, modelling the links between the state vector and the observations. Each equation can be inaccurate and a description of the errors is stated in terms of a Gaussian noise characterized by a covariance matrix. As it will be shown in the paper, Data Assimilation is a suitable approach to solve the three drawbacks of Weickert's methods because: the evolution model describes the images' temporal dynamics; the covariance associated to the observation equation error overcomes the problem of missing data by weighting, at each pixel, the contribution of the observation equation in the computation of the solution; Data Assimilation is a frame-by-frame process allowing an affordable computational cost.

## 2 The Data Assimilation framework

The Data Assimilation framework aims to solve the system (1,2,3) with respect to a state vector  $\mathbf{X}(\mathbf{x},t)$ , depending on the spatial coordinate  $\mathbf{x}$  and time  $t$ :

$$\frac{\partial \mathbf{X}}{\partial t}(\mathbf{x},t) + \mathbb{M}(\mathbf{X})(\mathbf{x},t) = \varepsilon_m(\mathbf{x},t) \quad (1)$$

$$\mathbf{X}(\mathbf{x},0) = \mathbf{X}_b(\mathbf{x}) + \varepsilon_b(\mathbf{x}) \quad (2)$$

$$\mathbb{H}(\mathbf{Y},\mathbf{X})(\mathbf{x},t) = \varepsilon_o(\mathbf{x},t) \quad (3)$$

Equation (1) describes the temporal evolution of  $\mathbf{X}$ .  $\mathbb{M}$ , called *evolution model*, is supposed differentiable. As  $\mathbb{M}$  approximately describes the evolution of the state vector, a *model error*  $\varepsilon_m$  is introduced to quantify inaccuracies. Equation (2) establishes the initial condition of the state vector and the error is expressed by the *background error*  $\varepsilon_b$ . Equation (3), called *observation equation*, describes the links, possibly complex and non-linear, between the observation vector,  $\mathbf{Y}(\mathbf{x},t)$ , and the state vector. The *observation error*  $\varepsilon_o$  represents the imperfection of  $\mathbb{H}$  and the measurement errors.  $\varepsilon_m$ ,  $\varepsilon_b$  and  $\varepsilon_o$  are assumed to be Gaussian and fully characterized by their covariance matrices  $Q$ ,  $B$  and  $R$ .

Let  $X$  denote a Gaussian stochastic vector depending on a space-time coordinate  $(\mathbf{x},t)$ .  $X = X(\mathbf{x},t)$  and

$X' = X(\mathbf{x}',t')$  are the values on two locations. The covariance matrix  $\Sigma$ , computed for  $X$  and  $X'$ , measures their dependency and is defined by:

$$\Sigma(\mathbf{x},t,\mathbf{x}',t') = \iint (X - \mathbb{E}X)^T (X' - \mathbb{E}X') dP_{X,X'}$$

$P_{X,X'}$  is the joint distribution of  $(X,X')$  and  $\mathbb{E}$  denotes the expectation.  $P_{X,X'}$  is a Gaussian distribution whose covariance matrix depends on  $\mathbf{x}$ ,  $t$ ,  $\mathbf{x}'$  and  $t'$ . The inverse of a covariance matrix is formally and implicitly defined (Oliver, 1998) by:

$$\iint \Sigma^{-1}(\mathbf{x},t,\mathbf{x}'',t'') \Sigma(\mathbf{x}'',t'',\mathbf{x}',t') d\mathbf{x}'' dt'' = \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') \quad (4)$$

with  $\delta$  denoting the Dirac function.

We note  $A = \Omega \times [0, \mathbf{T}]$ ,  $\Omega$  being the spatial domain and  $[0, \mathbf{T}]$  the temporal domain. In order to solve equations (1), (2) and (3), we define the functional  $E$ , to be minimized, as stated by equation (5), displayed at the top of next page. If  $\varepsilon_m$ ,  $\varepsilon_b$  and  $\varepsilon_o$  are assumed to be independent,  $E$  represents the log-likelihood of  $\mathbf{X}$ . The minimization is carried out using a calculus of variations: the differential of  $E$  with respect to  $\mathbf{X}$ , denoted  $\frac{\partial E}{\partial \mathbf{X}}$ , is calculated and the associated

Euler-Lagrange equation is solved.  $\frac{\partial E}{\partial \mathbf{X}}$  is obtained by (i) computing the derivative of  $E$  in the direction  $\eta$ :

$$\frac{\partial E}{\partial \mathbf{X}}(\eta) = \lim_{\gamma \rightarrow 0} \frac{d}{d\gamma} (E(\mathbf{X} + \gamma\eta)) \quad (6)$$

(ii) introducing an auxiliary variable  $\lambda$ , called *adjoint variable* in the literature:

$$\lambda(\mathbf{x},t) = \int_A Q^{-1}(\mathbf{x},t,\mathbf{x}',t') \left( \frac{\partial \mathbf{X}}{\partial t} + \mathbb{M}(\mathbf{X}) \right) (\mathbf{x}',t') d\mathbf{x}' dt' \quad (7)$$

and (iii) writing the state vector  $\mathbf{X}$  as  $\mathbf{X}_b + \delta\mathbf{X}$ . Vectors  $\mathbf{X}_b$  and  $\delta\mathbf{X}$  are respectively called *background vector* and *incremental vector* in the Data Assimilation literature. This leads to the following system of equations:

$$\lambda(\mathbf{x}, \mathbf{T}) = 0 \quad (8)$$

$$-\frac{\partial \lambda}{\partial t} + \left( \frac{\partial \mathbb{M}}{\partial \mathbf{X}} \Big|_{\mathbf{x}_b} \right)^* \lambda = - \int_A \left( \frac{\partial \mathbb{H}}{\partial \mathbf{X}} \Big|_{\mathbf{x}_b} \right)^* R^{-1} \quad (9)$$

$$\left( \mathbb{H}(\mathbf{X}_b, \mathbf{Y}) + \frac{\partial \mathbb{H}}{\partial \mathbf{X}} \Big|_{\mathbf{x}_b} (\delta\mathbf{X}) \right) d\mathbf{x}' dt'$$

$$\mathbf{X}_b(\mathbf{x},0) = \mathbf{X}_b(\mathbf{x}) \quad (10)$$

$$\frac{\partial \mathbf{X}_b}{\partial t} + \mathbb{M}(\mathbf{X}_b) = 0 \quad (11)$$

$$\delta\mathbf{X}(\mathbf{x},0) = \int_\Omega B\lambda(\mathbf{x}',0) d\mathbf{x}' \quad (12)$$

$$\frac{\partial \delta\mathbf{X}}{\partial t} + \frac{\partial \mathbb{M}}{\partial \mathbf{X}} \Big|_{\mathbf{x}_b} (\delta\mathbf{X}) = \int_A Q\lambda(\mathbf{x}',t') d\mathbf{x}' dt' \quad (13)$$

$$\begin{aligned}
E(\mathbf{X}) = & \int_A \int_A \left( \frac{\partial \mathbf{X}}{\partial t} + \mathbb{M}(\mathbf{X}) \right)^T (\mathbf{x}, t) Q^{-1}(\mathbf{x}, t, \mathbf{x}', t') \left( \frac{\partial \mathbf{X}}{\partial t} + \mathbb{M}(\mathbf{X}) \right) (\mathbf{x}', t') d\mathbf{x} dt d\mathbf{x}' dt' \\
& + \int_A \int_A \mathbb{H}(\mathbf{X}, \mathbf{Y})^T (\mathbf{x}, t) R^{-1}(\mathbf{x}, t, \mathbf{x}', t') \mathbb{H}(\mathbf{X}, \mathbf{Y}) (\mathbf{x}', t') d\mathbf{x} dt d\mathbf{x}' dt' \\
& + \int_{\Omega} \int_{\Omega} (\mathbf{X}(\mathbf{x}, 0) - \mathbf{X}_b(\mathbf{x}))^T B^{-1}(\mathbf{x}, \mathbf{x}') (\mathbf{X}(\mathbf{x}', 0) - \mathbf{X}_b(\mathbf{x}')) d\mathbf{x} d\mathbf{x}'
\end{aligned} \tag{5}$$

The reader is referred to (Valur Hólm, 2003) for details on the determination of the differential of  $E$  and Euler-Lagrange equations. The decomposition of the state vector into background and incremental vectors requires linearizing the operators  $\mathbb{M}$  and  $\mathbb{H}$  using a first-order Taylor development:  $\mathbb{M}(\mathbf{X}) \simeq \mathbb{M}(\mathbf{X}_b) + \left. \frac{\partial \mathbb{M}}{\partial \mathbf{X}} \right|_{\mathbf{X}_b} (\delta \mathbf{X})$ . The operator  $\left( \frac{\partial \cdot}{\partial \mathbf{X}} \right)^*$  is called *adjoint operator* and represents a compact notation for integration by parts of the differential of the operators  $\mathbb{M}$  and  $\mathbb{H}$ . It is formally defined by:

$$\left\langle \frac{\partial \mathbb{M}}{\partial \mathbf{X}}(\eta), \lambda \right\rangle = \left\langle \eta, \left( \frac{\partial \mathbb{M}}{\partial \mathbf{X}} \right)^*(\lambda) \right\rangle \tag{14}$$

Riesz's theorem ensures the existence and uniqueness of the adjoint operator. To solve the system (8) to (13), equation (11) is first solved using initial condition (10) and provides  $\mathbf{X}_b$ . Equation (9) is solved retrogradely to provide  $\lambda$  using initial condition (8). The incremental vector is finally provided by solving (13) with (12) as initial condition. If  $\mathbb{M}$  and  $\mathbb{H}$  are not linear, an iterative algorithm is used to sequentially improve the solution of (8,9,12,13): at each iteration, the value  $\delta \mathbf{X}$  is computed from the background vector  $\mathbf{X}_b$  and the value  $\delta \mathbf{X} + \mathbf{X}_b$  becomes the background vector for the next iteration.

### 3 Image Assimilation

This section explains how to solve ill-posed Image Processing problems using the framework of Data Assimilation and constitutes the core of the research and the main contribution of the paper. Using Data Assimilation to solve Image Processing Problem is an emerging domain. Studies have been done about curve tracking (Papadakis and Mémin, 2007) and determination of optical flow (Papadakis et al., 2007a; Papadakis et al., 2007b). In (Huot et al., 2008), a method is proposed to estimate the ocean surface circulation from SST images by assimilating image data in an image model for producing pseudo-observations for the oceanographic model. In this paper, we consider the general case of ill-posed problems, solved using Tikhonov regularization methods, and we define a method to convert them, in a generic way, from Tikhonov regularization to Data Assimilation framework.

First, state and observation vectors have to be defined. Obviously, the observations will be images or processed images, and the content of the state vector will strongly depends on the studied application. For example, segmentation, denoising and restoration use a state vector which is an image. Tracking, image registration and motion estimation are represented by a vector field. Active contours are modelled by curves. Sometimes, the links between the state vector and the observation are highly indirect. For instance, as it is not possible to deduce the ocean circulation from surface temperatures by a shallow-water model, an additional system has to be considered with relevant evolution and observation equations (Huot et al., 2008).

To solve an Image Processing problem with the 4D-Var algorithm, as described in Section 2, a suitable evolution model describing the temporal evolution of the state vector has to be stated. Its associated error must also be defined. Next, an observation equation and its error should characterize the properties of observations. Last, the initial condition and its background error have to be established. This is discussed in the next Subsections.

#### 3.1 The evolution model and error

One first choice of evolution model is to consider the simple transport equation  $\frac{d\mathbf{X}}{dt} = 0$ , or:

$$\frac{\partial \mathbf{X}}{\partial t} + \frac{\partial \mathbf{X}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial t} = 0 \tag{15}$$

The evolution model is then obtained by identification:  $\mathbb{M}(\mathbf{X})(\mathbf{x}, t) = \frac{\partial \mathbf{X}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial t}$ . Section 4 describes the use of equation (15) for determining optical flow.

Another possible choice is to express the transport of the state vector as a diffusion, a physical law ruling the transport of chemical species or temperature.

The general formulation is:  $\frac{\partial \mathbf{X}}{\partial t} = \nabla^T (D \nabla \mathbf{X})$  with

$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T$ .  $D$  is a tensor characterizing the direction and intensity of the diffusion. It is possible to drive this diffusion according to the image characteristics. A well-known example is the Perona & Malik diffusion: the tensor matrix is equal to  $c(\|\nabla \mathbf{X}\|) Id$  with  $c$  a Gaussian function and  $Id$  the identity matrix.

One property of this diffusion is to smooth the image while preserving the contours. In this case, identification gives:  $\mathbb{M}(\mathbf{X}) = -\nabla^T(D\nabla\mathbf{X})$ .

The matrix  $Q$  of the model error  $\epsilon_m$  has to be specified. It is first important to understand the action of  $Q^{-1}$  in the functional (5). Let us describe two possible choices for  $Q$  and analyze their impact in a functional  $E(\mathbf{X}) = \iint F^T(\mathbf{X})Q^{-1}F(\mathbf{X})d\mathbf{x}d\mathbf{x}'$  which has to be minimized.

First, let  $Q$  being the Dirac covariance defined by  $Q(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}')$ . We have  $E(\mathbf{X}) = \int_{\Omega} \|F(\mathbf{X})\|^2 d\mathbf{x}$ : a Dirac covariance is corresponding to a zero-order regularization.

A second choice is the exponential covariance defined by  $Q(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\mathbf{x} - \mathbf{x}'}{\sigma}\right)$ . We have  $E(\mathbf{X}) = \frac{1}{2\sigma} \int \left( \|F(\mathbf{X})\|^2 + \sigma^2 \left\| \frac{\partial F(\mathbf{X})}{\partial \mathbf{x}} \right\|^2 \right) d\mathbf{x}$ . The exponential covariance is associated to a first-order regularization.

In the general case, the inversion of a covariance matrix is often non-trivial and usually inaccessible. Restrictive choices have to be made such as the previous ones. For further details, the reader is referred to (Oliver, 1998; Tarantola, 2005; Béréziat and Herlin, 2008).

### 3.2 The observation equation and error

As previously pointed out in Section 2, the observation equation describes the links between the state vector and the observations. As the regularization part of the Tikhonov's method does not contain observation values, the observation equation will be exactly corresponding to the Image Model existing in the Tikhonov's approach.

The observation error has to be stated and its covariance matrix  $R$  specified.  $R$  has then to be inverted (see equation (9)) and, as explained, this is only possible for simple cases. The Dirac covariance can be used to express a null interaction between two space-time locations. This allows observation errors to be located and quantified in space and time.  $R$  is written as:

$$R(\mathbf{x}, t, \mathbf{x}', t') = r(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') \quad (16)$$

with  $r$  a real invertible matrix and:

$$R^{-1}(\mathbf{x}, t, \mathbf{x}', t') = \delta(\mathbf{x} - \mathbf{x}') \delta(t - t') r^{-1}(\mathbf{x}, t) \quad (17)$$

The matrix  $r^{-1}$  characterizes the quality of the observation: a high value indicates that the observation is relevant. Assuming the availability of a function  $f$  measuring the confidence in observation values

( $f \in [0, 1]$ ,  $f = 0$  for no confidence),  $r^{-1}$  can be defined by:

$$r^{-1}(\mathbf{x}, t) = r_0(1 - f(\mathbf{x}, t)) + r_1 f(\mathbf{x}, t) \quad (18)$$

$r^{-1}(\mathbf{x}, t)$  will be close to a "minimal value"  $r_0$  on regions where confidence is low and  $r_1$  otherwise.  $r_0$  and  $r_1$  are constant and invertible matrices. For efficiency,  $f$  is modeled as:

$$f(\mathbf{x}, t) = f_{\text{noise}}(\mathbf{x}, t) f_{\text{sensor}}(\mathbf{x}, t) f_{\mathbb{H}}(\mathbf{x}, t) \quad (19)$$

$f_{\text{noise}}$  characterizes noise acquisition: it is close to 0 for large noise.  $f_{\text{sensor}}$  is equal to 0 when data are not acquired by the sensor.  $f_{\mathbb{H}}$  measures the confidence in the observation model; it is close to 0 if the observation equation is not valid.

Observation values with a low confidence will then not be taken into account for the computation of the state vector and for the solution of the Image Processing problem.

## 4 Determining optical flow

Let  $I$  be a sequence of images on a spatial domain  $\Omega$  and  $\mathbf{W}(\mathbf{x}, t)$  be the velocity vector of a point  $\mathbf{x} \in \Omega$  at time  $t$ . The velocity vector is expressed by the Optical Flow Constraint equation (Horn and Schunk, 1981):

$$\nabla I^T(\mathbf{x}, t) \mathbf{W}(\mathbf{x}, t) + \frac{\partial I}{\partial t}(\mathbf{x}, t) = 0 \quad \forall \mathbf{x} \in \Omega \quad (20)$$

This is an ill-posed problem: the velocity vector has two components and the optical flow equation is not sufficient to compute both. A solution can be obtained using Tikhonov's method by constraining spatially (Horn and Schunk, 1981) or spatially and temporally (Weickert and Schnörr, 2001) the solution. Let us explain, advantages of the data assimilation methods on this simple example.

### 4.1 Observation and evolution models

$\mathbf{W}(\mathbf{x}, t)$  is considered as the state vector  $\mathbf{X}(\mathbf{x}, t)$  and the image gradients  $(\nabla I(\mathbf{x}, t), I_t(\mathbf{x}, t))$  constitutes the observation vector  $\mathbf{Y}(\mathbf{x}, t)$ . By identification in (3) and (20), the observation model is:

$$\mathbb{H}(\mathbf{W}, I) = \nabla I(\mathbf{x}, t)^T \mathbf{W}(\mathbf{x}, t) + I_t(\mathbf{x}, t) \quad (21)$$

To determine a suitable observation error covariance handling missing data, we use equations (17), (18) and (19) to define the inverse covariance.  $f_{\text{noise}}$  is assumed to be equal to 1 without further information.  $f_{\text{sensor}}$  is set to 0 where data are not acquired and to 1 otherwise.  $f_{\mathbb{H}}$  is chosen from the following remark: on a sequence of images with regions

of uniform grey level values, the spatio-temporal gradient is null on these regions and equation (20) is degenerated. For avoiding considering these points,  $f_{\mathbb{H}}$  is:

$$f_{\mathbb{H}}(\mathbf{x}, t) = 1 - \exp(-\|\nabla_3 I(\mathbf{x}, t)\|^2) \quad (22)$$

where  $\nabla_3$  denotes the spatio-temporal gradient. The observation model being scalar,  $r_0$  and  $r_1$  are scalar and respectively set to  $\varepsilon$  and  $1 - \varepsilon$  with  $\varepsilon \sim 10^{-6}$ .

The transport of the velocity is chosen as evolution equation. The evolution model is:

$$\begin{aligned} \mathbb{M}(\mathbf{W}) &= (\mathbb{M}_1(\mathbf{W}) \quad \mathbb{M}_2(\mathbf{W})) \\ &= \begin{pmatrix} UU_x + VU_y \\ UV_x + VV_y \end{pmatrix}^T \end{aligned}$$

with  $(U, V)^T = \mathbf{W}$ .

Let us denote  $I_2$  the  $2 \times 2$  identity matrix.  $Q$  is chosen as:

$$Q(\mathbf{x}, t, \mathbf{x}', t') = I_2 \exp\left(-\frac{1}{\sigma}(\|\mathbf{x} - \mathbf{x}'\| + |t - t'|)\right) \quad (23)$$

to ensure a first-order spatial regularization.

We choose the Horn and Schunk's algorithm to compute the velocity field on the first image of the sequence. This initial condition is supposed to be without errors:  $B(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}')$ .

## 4.2 Adjoint operators

In order to determine the adjoint operators of  $\mathbb{M}$  and  $\mathbb{H}$ , the directional derivatives must first be established.

Using the definition (6), we obtain:

$$\begin{aligned} \frac{\partial \mathbb{M}_1}{\partial \mathbf{W}}(\eta) &= \frac{\partial \mathbb{M}_1}{\partial U}(\eta^1) + \frac{\partial \mathbb{M}_1}{\partial V}(\eta^2) \\ &= U\eta_x^1 + U_x\eta^1 + V\eta_y^1 + U_y\eta^2 \\ \frac{\partial \mathbb{M}_2}{\partial \mathbf{W}}(\eta) &= \frac{\partial \mathbb{M}_2}{\partial U}(\eta^1) + \frac{\partial \mathbb{M}_2}{\partial V}(\eta^2) \\ &= U\eta_x^2 + V_y\eta^2 + V\eta_y^2 + V_x\eta^1 \end{aligned}$$

with  $\eta = (\eta^1 \quad \eta^2)^T$ . By definition (14) and considering border terms equal to zero, the adjoint operator of  $\mathbb{M}$  is :

$$\begin{aligned} \left(\frac{\partial \mathbb{M}_1}{\partial \mathbf{W}}\right)^*(\lambda) &= -U\lambda_x^1 - V_y\lambda^1 - V\lambda_y^1 + U_y\lambda^2 \\ \left(\frac{\partial \mathbb{M}_2}{\partial \mathbf{W}}\right)^*(\lambda) &= -U_x\lambda^2 - U\lambda_x^2 - V\lambda_y^2 + V_x\lambda^1 \end{aligned}$$

with  $\lambda = (\lambda^1 \quad \lambda^2)^T$ .

The directional derivative of the observation operator is:

$$\frac{\partial \mathbb{H}}{\partial \mathbf{W}}(\eta)(\mathbf{x}, t) = \nabla I^T(\mathbf{x}, t)\eta(\mathbf{x}, t)$$

and the adjoint operator is:

$$\left(\frac{\partial \mathbb{H}}{\partial \mathbf{W}}\right)^*(\lambda)(\mathbf{x}, t) = \nabla I(\mathbf{x}, t)\lambda(\mathbf{x}, t)$$

## 4.3 Discretization

Using the choices explained in Subsection 4.1, the three PDEs (9,11,13) become:

$$\frac{\partial \mathbf{W}}{\partial t} + \mathbf{W}^T \nabla \mathbf{W} = 0 \quad (24)$$

$$-\frac{\partial \lambda}{\partial t} - \nabla \lambda^T \mathbf{W} - \lambda^T \nabla^\perp \mathbf{W} = A * L \quad (25)$$

$$\frac{\partial \delta \mathbf{W}}{\partial t} + \mathbf{W}^T \nabla \delta \mathbf{W} + \nabla \mathbf{W}^T \delta \mathbf{W} = Q * \lambda \quad (26)$$

with:  $A = -\frac{1}{2\sigma^2} \nabla I R^{-1}$ ,  $L = I_t + \nabla I^T(\mathbf{W} + \delta \mathbf{W})$ ,  $\nabla^\perp \mathbf{W} = (\nabla^\perp U \quad \nabla^\perp V)$  and  $\nabla^\perp U = (U_y \quad -U_x)^T$ . These three equations are further approximated using the finite difference technique.

As  $\mathbf{W}(\mathbf{x}, t)$  is a vector of  $\mathbb{R}^2$ , equation (24) has two components. The first one combines a linear advection in direction  $y$  and a non linear in direction  $x$ . A direct approximation leads to an unstable numerical scheme. In (Papadakis et al., 2007a), the scheme is stabilized by introducing a diffusive term with the drawback to smooth the solution. We propose an alternative to stabilize the numerical scheme associated to the original equation, as explain in the following. The first component of equation (24) is first expressed as a two-equation system using a splitting method (Verwer and Sportisse, 1998):

$$\frac{\partial U}{\partial t} + UU_x = 0 \quad (27)$$

$$\frac{\partial U}{\partial t} + VU_y = 0 \quad (28)$$

Equation (27) is rewritten using the Lax-Friedrich method (Sethian, 1996) in the following form  $\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} = 0$  with  $F(u) = \frac{1}{2}u^2$ . This new equation is approximated by the explicit scheme:

$$U_{i,j}^{k+1} = \frac{1}{2}(U_{i+1,j}^k + U_{i-1,j}^k) - \frac{\Delta t}{2}(F_{i+1,j}^k - F_{i-1,j}^k)$$

with  $U_{i,j}^k = U(x_i, y_i, t_k)$ ,  $F_{i,j}^k = F(U(x_i, y_i, t_k))$ . The term  $\frac{1}{2}(U_{i+1,j}^k + U_{i-1,j}^k)$  stabilizes the numerical scheme (by adding a diffusive effect) when  $\Delta t$  has a low value (the Courant-Friedrich-Levy condition). The linear advection (28) is approximated using a shock explicit scheme (Sethian, 1996):

$$U_{i,j}^{k+1} = U_{i,j}^k - \Delta t S_x(U^k, V^k)_{i,j}$$

with  $S$  the shock operator, defined in the  $x$ -direction by:  $S_x(U, V)_{i,j} = \max(V_{i,j}, 0)(U_{i,j} - U_{i-1,j}) + \min(V_{i,j}, 0)(U_{i+1,j} - U_{i,j})$ .

It can be seen that the second component of (24) contains a linear advection term in direction  $x$  and a non linear advection term in direction  $y$ . The same strategy is then used to perform the discretization process.

Equation (25) combines a linear advection, a term of reaction and a forcing term. It also has two components. The first is  $-\frac{\partial \lambda^1}{\partial t} - U\lambda_x^1 - V_y\lambda^1 - V\lambda_y^1 + U_y\lambda^2 = A^1 \star L$  with  $A^1 = -\frac{1}{2\sigma^2}I_x R^{-1}$ . It is split into two parts. The first one contains the linear advection in direction  $x$  and the reaction term  $-\frac{\partial \lambda^1}{\partial t} - U\lambda_x^1 - V_y\lambda^1 = 0$  and is approximated in the same way as (28) with an explicit shock scheme. However, the equation is retrograde as the initial condition is given at time  $T$ :

$$(\lambda^1)_{i,j}^{k-1} = \left(1 + \frac{1}{2}(V_{i,j+1}^k - V_{i,j-1}^k)\right) (\lambda^1)_{i,j}^k + \Delta t S_x((\lambda^1)^k, U^k)$$

The second part contains the linear advection term in direction  $y$  and the forcing term:  $-\frac{\partial \lambda^1}{\partial t} - V\lambda_y^1 + U_y\lambda^2 = A^1 \star L$ . Again, a shock explicit scheme is used:

$$(\lambda^1)_{i,j}^{k-1} = (\lambda^1)_{i,j}^k - \frac{\Delta t}{2} (U_{i,j+1}^k - U_{i,j-1}^k) (\lambda^2)_{i,j}^k - \Delta t (A^1 \star L)_{i,j}^k - \Delta t S_y((\lambda^1)^k, V)$$

Having the same structure as the first component, the second component of (25) is approximated using the same method

The last equation, (26), is similar to equation (25) and the same discretization technique is applied.

#### 4.4 Results

The “taxi” sequence and a synthetic sequence have been chosen for discussing results. For both cases, the initial condition is provided by the Horn and Schunk method applied on the two-first frames of the sequence. Image gradients, providing observations, are computed with a convolution method and a derivative Gaussian kernel. The parameter  $\sigma$  of the matrix  $Q$  is set to 1. The number of iterations is set to 5.

The taxi sequence displays several cars with a slow and quasi uniform motion.

In a first experiment, we compute the optical flow on the sequence using the Data Assimilation method (D.A.). The Horn & Schunk method (H.&S.) is also applied and results, on the fifth frame, are displayed



Figure 1: Result on frame 5: D.A. (top), H&S (bottom).

for comparison purposes on Figure 1. Results of both methods are similar in this case because observation values are available on the whole sequence obviously. Having chosen H.&S. or another method does not really matter. The issue is not comparing optical flow method but proving the efficiency of D.A. for missing data or complex dynamics. Then, a second experiment is designed to prove the capability of D.A. to manage missing data. To simulate a sensor failure, a large region around the white car is set to zero in a frame and the function  $f_{\text{sensor}}$  returns zero on these pixels. Image gradients are then computed. Figure 2 shows the results for D.A. and H.&S. methods. This latter obviously fails to provide acceptable velocity vectors over the missing data region while D.A. succeeds due to the eviction of aberrant values in the computation. Even a whole frame of observation can be missing: we force image gradient values to zero on the fifth frame of the taxi sequence and  $f_{\text{H}} \equiv 0$  on this frame. Figure 3 displays results obtained with D.A. with and without image gradients observation on frame 5. Results are similar, due to the fact that the evolution model correctly describes the temporal dynamics of the sequence and compensate missing data.



Figure 2: Missing data: D.A. (top), H.&S. (bottom).



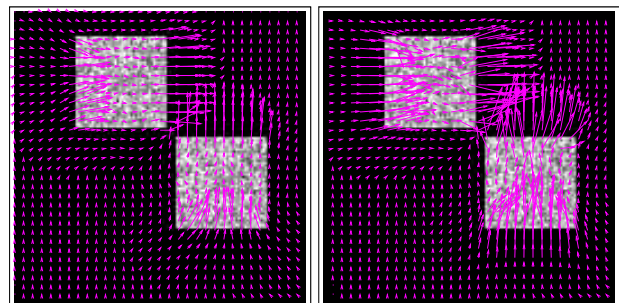
Figure 3: Result with gradients set to zero on frame 5 (top) and with gradient available (bottom).

H.&S. method is no more able to provide any result in that case, conducting to a discontinuity in the output.

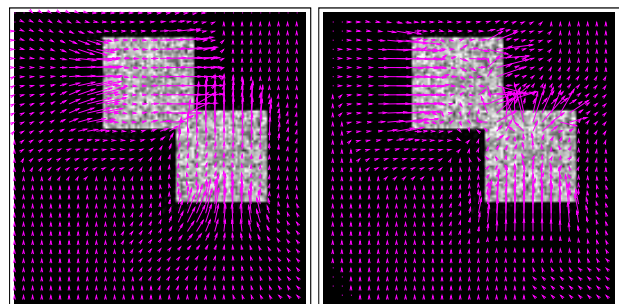
A third experiment is dedicated to the demonstration that the evolution model improves the estimation of the image dynamics. For that purpose, we build a synthetic sequence with two squares, one moving horizontally and one vertically. At the end of the sequence, they meet each other. Figure 4 shows the result of both methods at the beginning and the end of the sequence. H.&S.'s method fails to estimate a correct velocity direction when the squares meet. This is a common drawback of image processing method due to over regularization. D.A. however provides accurate results: the evolution model correctly describes the squares dynamics and decreases the effect of spatial regularization.

## 5 Conclusion

In this paper we described a framework to solve ill-posed Image Processing problems using Data Assimilation algorithms. This approach is an alternative to the Weickert's method, which constrains the solu-



(a) frame 4



(b) frame 10

Figure 4: Result with D.A. (left) and H.&S. (right).



tion's variations in space and time.

General principles to choose a suitable evolution model, an observation equation, and their matrices of covariance were discussed in the paper. The observation equation corresponds to the Image Model of the Tikhonov regularization method. The evolution equation constrains the solution in time and the model error constrains it in space. As an illustration, we proved that an exponential covariance model error is equivalent to a first order Tikhonov regularization. The three drawbacks of the Weickert's method, pointed out in the Introduction, are overcome. First, the state vector's dynamics is described by a suitable evolution model. Second, a covariance  $R$  built from the confidence on the observation values weights the influence of these latter: observations having a low quality, such as missing data, are not taken into account in the computation of the solution. And last, the algorithm processes a sequence of images frame by frame, which allows an implementation with low memory requirements. The choice of the evolution model remains a key point as it greatly depends on the applicative context. We described it in the case of determining optical flow: image's grey level values are assimilated in an evolution model representing a transport equation of the velocity and a stable numerical scheme, approximating this equation, has been established.

The main perspective of the work is to define a generic form of the evolution model which will then be specified depending of the Image Processing application. Diffusion laws, modelling the transport of the state vector using the physical conservative principles, are good candidates for a number of reasons. First, their approximation in a discrete space is robust as stable numerical schemes exist. Second, the adjoint of the model operator is known explicitly. Third, several types of diffusion can be considered depending on their impact on the spatial and temporal regularization of the solution. An isotropic diffusion is equivalent to a uniform regularization but has the drawback to smooth the solution while an adaptive diffusion, driven by the state vector, allows discontinuities to be preserved. As the spatial regularization may also be performed by the covariance matrix  $Q$ , it could be interesting to compare the two paradigms. Moreover, the diffusion can also be driven by the observation values or by their confidence. For instance, a strong diffusion on the state vector may operate when the observation confidence is low which is an alternative way to deal with missing data. Again, a comparison between such a diffusion and the covariance matrix  $R$  built from the observation confidence must be led.

## References

- Béréziat, D. and Herlin, I. (2008). Solving ill-posed image processing problems using data assimilation. Application to optical flow. Research Report 6477, INRIA.
- Hadamard, J. (1923). *Lecture on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press, New Haven.
- Horn, B. and Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.
- Huot, E., Herlin, I., and Korotaev, G. (2008). Assimilation of sst satellite images for estimation of ocean circulation velocity. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, Massachusetts, U.S.A.
- Oliver, D. (1998). Calculation of the inverse of the covariance. *Mathematical Geology*, 30(7):911–933.
- Papadakis, N., Corpetti, T., and Mémin, E. (2007a). Dynamically consistent optical flow estimation. In *Proceedings of International Conference on Computer Vision*, Rio de Janeiro, Brazil.
- Papadakis, N., Héas, P., and Mémin, E. (2007b). Image assimilation for motion estimation of atmospheric layers with shallow-water model. In *Proceedings of Asian Conference on Computer Vision*, pages 864–874, Tokyo, Japan.
- Papadakis, N. and Mémin, E. (2007). Variational optimal control technique for the tracking of deformable objects. In *Proceedings of International Conference on Computer Vision*, Rio de Janeiro, Brazil.
- Sethian, J. (1996). *Level Set Methods*. Cambridge University Press.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.
- Tikhonov, A. N. (1963). Regularization of incorrectly posed problems. *Sov. Math. Dokl.*, 4:1624–1627.
- Valur Hólm, E. (2003). Lectures notes on assimilation algorithms. Technical report, European Centre for Medium-Range Weather Forecasts Reading, U.K.
- Verwer, J. and Sportisse, B. (1998). A note on operator splitting in a stiff linear case. Technical Report MAS-R9830, Center voor Wiskunde en Informatica.
- Weickert, J. and Schnörr, C. (2001). Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14:245–255.