# OPTIMAL CONTROL THEORY FOR MULTI-RESOLUTION PROBLEMS IN COMPUTER VISION
## Application to Optical-flow Estimation

Pascal Zille[1,2] and Thomas Corpetti[2]

[1]*UMR CNRS LMFA, Ecole Centrale Lyon, Lyon, France*
[2]*CNRS LIAMA TIPE, Beijing, China*

Keywords:     Multi-resolution, Optical-flow, Optimal Control, Data Assimilation.

Abstract:     This paper is concerned with the multi-resolution issue used in many computer vision applications. Such approaches are very popular to optimize a cost function that, in most of the situations, has been linearized for mathematical facility reasons. In general, a multi-resolution setup consists in a redefinition of the problem at a different resolution level where the mathematical assumptions (usually linearity) hold. Following a coarse-to-fine strategy, a usual process consists in 1) optimizing the large scales and 2) use this result as an initial condition for the estimation at finer scales. Such process is repeated until the plain image resolution. One of the main drawbacks of such downscaling approach is its incapacity to correct the eventual errors that have been made at larger scales. These latter are indeed propagated along the scales and disturb the final result. In this paper, we suggest a new formulation of the multi-resolution setup where we exploit some smoothing techniques issued from optimal control theory and in particular variational data assimilation. The time is here artificial and is related to the various scales we are dealing with. Following a consistent mathematical framework, we define an original downscaling/upscaling technique to perform the multi-resolution. We validate this approach by defining a simple optical flow estimation technique based on Lucas-Kanade. Experimental results on synthetic data demonstrate the efficiency of this new methodology.

# 1 INTRODUCTION

A number of common computer vision techniques (related to motion estimation, segmentation, characterization, ...) are defined as finding a variable $X$ by solving an equation $\mathcal{H}(X, I, \boldsymbol{X}^0)$ that depends on the image luminance $I$ and the pixel grid $\boldsymbol{X}^0$ (also noted $\boldsymbol{x}$ in this paper). Because most of the usual assumptions hold only in a linear case for obvious mathematical properties, it is common to embed the resolution of $\mathcal{H}(X, I, \boldsymbol{X}^0)$ in a so-called "multi-resolution" scheme (Burt, 1988; Mallat, 1989). The main principle consists in redefining the images on smaller grids $\boldsymbol{X}^N$ that correspond to the initial grid $\boldsymbol{X}^0$ divided by a factor $N$. On such "coarse" images, we assume that de resolution of $\mathcal{H}(X, I, \boldsymbol{X}^N)$ can be done under linear constraints and this provide a coarse approximation $X^N$ of the final solution $X$. In a step forward, this approximation $X^N$ is used as an initial condition for a new problem where the goal is to extract a refinement $dX$ (with $X = X^N + dX$) that can be estimated from $X^N$ using linear constraints. Such process is commonly

repeated for several levels of resolutions, yielding a succession of linear problems.

This is for instance a usual way to deal with optical-flow estimation where the common brightness consistency assumption (that links the spatial and temporal gradients of the luminance with the spatial velocity $\boldsymbol{v}(\boldsymbol{x})$ to estimate), defined as:

$$\frac{\partial I}{\partial t} + \boldsymbol{v}(\boldsymbol{x}) \cdot \nabla I \approx 0 \qquad (1)$$

holds only for small displacements. When embedded in a multi-resolution scheme, this relation holds for the images at the coarser resolutions and the associated estimations are used as initial solutions for the finer resolutions.

To get such "coarse" data on which the problem is solved, a usual strategy consists in using a pyramidal decomposition of the images, for instance with wavelet decompositions or gaussian filtering followed by decimations, as illustrated in FIG. 1(left) for a factor $N = 2$. The estimation is then performed first from the smallest images (under linear constraints) and the
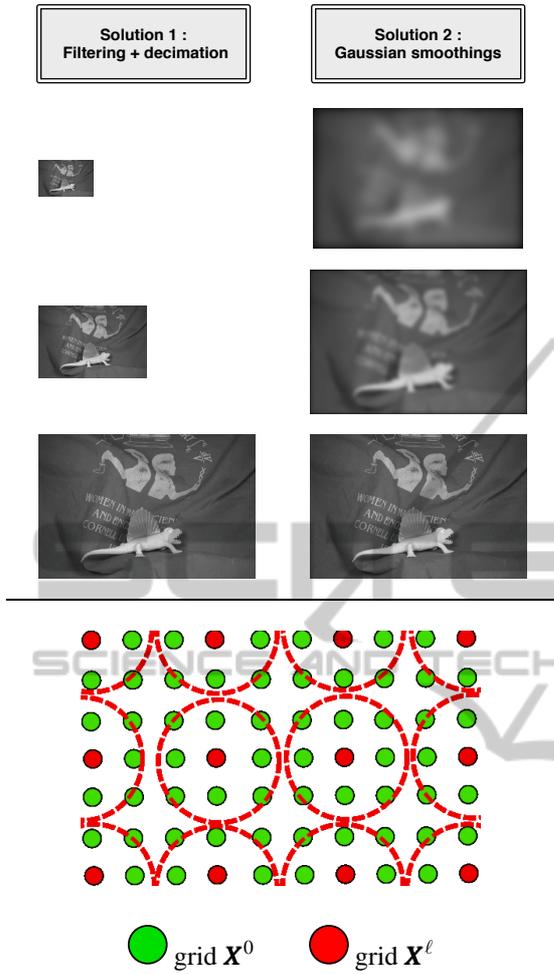
Figure 1: **Multiresolution strategies**. Top left: classical pyramidal decomposition; top right: successive gaussian convolutions. Bottom: illustration of the latter technique. Green circles represent the original pixel grid whereas the red ones represent the grid at a given resolution $\ell$.

solution is reprojected to the following level. Another possibility proposed in (Corpetti and Mémin, 2012) consists in obtaining $X^\ell$ (related to the solution of a problem $\mathcal{H}(X,I,\mathbf{X}^\ell)$ at a resolution $\ell$) by performing a convolution of $\mathcal{H}(X,I,\mathbf{X}^0)$ by a gaussian kernel of standard deviation $\ell$. Indeed, a multi-resolution scheme consists in redefining the problem on a grid $\mathbf{X}^\ell$ which can be viewed as a coarse representation of the initial grid $\mathbf{X}^0 = \mathbf{X}$ with a Brownian isotropic uncertainty of constant variance $\ell$ :

$$\mathbf{X}^\ell = \mathbf{X}^0 + \ell\, \mathbb{I}_2 d\mathbf{B}, \qquad (2)$$

where $\mathbf{B}$ is a standard Brownian motion and $\mathbb{I}$ the 2D identity matrix (see the illustration in figure 1). Under this scheme, any solution $X^\ell$ of a problem $\mathcal{H}(X,I,\mathbf{X}^\ell)$ defined on a grid $\mathbf{X}^\ell$ should satisfy the expectation $E(\mathcal{H}(X,I,\mathbf{X}^\ell)|\mathbf{X}^0)$ which is equivalent

(see (Corpetti and Mémin, 2012) for the demonstration) to a convolution of $\mathcal{H}(X,I,\mathbf{X}^0)$ with the isotropic centered gaussian $\mathcal{N}(0,\ell)$ of variance $\ell$. Therefore the multi-resolution can be performed by solving a family of problems $g_\ell * \mathcal{H}(X,I,\mathbf{X}^0)$ at the various resolutions $\ell$. A main advantage of such a formulation of the multi-resolution setup is to naturally get rid of the intrinsic problems related to pyramidal image decompositions (decimations and interpolations in particular). In addition, instead of dealing with successive decimations of factor 2 of the initial image to fix the different multiresolution levels, the evolutions of the levels $\ell$ are much flexible here.

**General difficulties of multi-resolution systems**: even if several specific techniques have been proposed for some applications (see for instance (Alvarez et al., 2000; Baatz and Schaape, 2000; Bajcsy and Kovacic, 1989; Brox et al., 2004; Eck et al., 1995; Ojala et al., 2002)), in general, whatever the multi-resolution setup chosen, one of the main difficulty remains on the succession of *independent* problems: at a given resolution $\ell_n$, the problem consists in finding $dX^{\ell_n}$ using $X^{\ell_{n+1}}$ as a coarse approximation: $X^{\ell_n} = X^{\ell_{n+1}} + dX^{\ell_n}$. Once $X^{\ell_n}$ estimated, its value is kept during the rest of the process. This is somewhat prejudicial since it is now recognized that small scales (related to finer resolutions) interact with larger scales. With such schemes, at a given resolution $\ell_n$, the information related to smaller scales $\ell < \ell_n$ can not be taken into account. In addition, any error in the estimation of $X^{\ell_n}$ will also be kept and propagated across the resolutions without any possibilities of correction.

To deal with these difficulties, we propose in this paper an original solution based on optimal control theory and in particular on data assimilation. Such schemes consist in estimating a sequence of unknowns $X(t)$ driven by a more or less exact dynamical model by performing a series of forward/backward integrations. The key idea of this study is to exploit such framework where the time will in fact be related to the various scales. The forward/backward integrations will then lead to a set of upscaling/downscaling approaches that appear to be adapted to our problem. Before entering into the core of the technique in section 3, we first briefly introduce the data assimilation framework.

## 2 DATA ASSIMILATION

In this section we will describe the general principles

of the assimilation scheme we devised in this study. Variational data assimilation is a technique derived from optimal control theory (Lions, 1971) to recover a state variable's trajectory from a sequence of measurements. Opposite to sequential Bayesian filters, which share the same aim, this framework allows to handle high dimensional systems (and is thus intensively used for instance in environmental sciences (Bennett, 1992; Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987)). We refer the reader to (Bennett, 1992; Le Dimet and Talagrand, 1986; Lions, 1971; Talagrand and Courtier, 1987; Talagrand, 1997; Vidard et al., 2000) for complete methodological aspects of data assimilation and applications concerning geophysical flows.

The problem consists in recovering, from an initial condition $X_0$, a system's state $X$ partially observed and driven by approximately known dynamics. This can be formalized as finding $X(\boldsymbol{x}, t)$, for any location $\boldsymbol{x} \in \Omega$ at time $t \in [t_0, t_f]$, that satisfies the system:

$$\frac{\partial X}{\partial t}(\boldsymbol{x}, t) + \mathbb{M}(X(\boldsymbol{x}, t)) = \nu_m(\boldsymbol{x}), \qquad (3)$$

$$X(\boldsymbol{x}, t_0) = X_0(\boldsymbol{x}) + \nu_n(\boldsymbol{x}), \qquad (4)$$

$$\boldsymbol{Y}(\boldsymbol{x}, t) = \mathbb{H}(X(\boldsymbol{x}, t)) + \nu_o(\boldsymbol{x}, t), \qquad (5)$$

where $\mathbb{M}$ is the non-linear operator relative to the dynamics, $X_0$ is the initial vector at time $t_0$ and $(\nu_n, \nu_m)$ are (unknown) additive control variables relative to noise on the dynamics and the initial condition respectively. In addition, noisy measurements $\boldsymbol{Y}$ of the unknown state are available through the non-linear operator $\mathbb{H}$, up to $\nu_o$. To estimate the system's state, a common methodology relies on the minimization of the cost function $\mathcal{J}$:

$$\begin{aligned} \mathcal{J}(X) = &\frac{1}{2} \int_{t_0}^{t_f} \|\boldsymbol{Y} - \mathbb{H}(X(\nu_m, \nu_n))\|_{R^{-1}}^2 dt \\ &+ \frac{1}{2} \|X(\boldsymbol{x}, t_0) - X_0(\boldsymbol{x})\|_{B^{-1}}^2 \qquad (6) \\ &+ \frac{1}{2} \int_{t_0}^{t_f} \|\frac{\partial X}{\partial t}(\boldsymbol{x}, t) + \mathbb{M}(X(\boldsymbol{x}, t))\|_{Q^{-1}}^2 dt, \end{aligned}$$

where we have introduced the information matrices $R, B, Q$ relative to the covariance of the errors $(\nu_m, \nu_n, \nu_o)$. The Mahalanobis distance that has been used reads, for an information matrix $A$: $\|X\|_{A^{-1}} = X^T A^{-1} X$. The evaluation of $X$ can be done by canceling the gradient $\delta \mathcal{J}_X(\theta) = \lim_{\beta \to 0} \frac{\mathcal{J}(X + \beta\theta) - \mathcal{J}(X)}{\beta}$ of (6). Unfortunately, the estimation of such gradient is in practice unfeasible for a large system's state since it would be necessary to integrate the dynamical model along all possible perturbations of the components of $X$. This is computationally impossible with actual hardwares when one deals with a complete sequence of images and complex dynamical models. One way

to cope with this difficulty is to write *an adjoint formulation* of the problem. To that end, the *adjoint variables* $\boldsymbol{\lambda}$ that express the errors of the dynamic model are introduced as:

$$\boldsymbol{\lambda}(\boldsymbol{x}, t) = \int_{\Omega, t} Q^{-1} \left( \frac{\partial X}{\partial t} + \mathbb{M}(X) \right) dx' dt' \qquad (7)$$

Denoting:

- $\left( \frac{\partial \mathbb{M}}{\partial \widetilde{X}} \right)$ and $\left( \frac{\partial \mathbb{H}}{\partial \widetilde{X}} \right)$ the *linear tangent operators* of $\mathbb{M}$ and $\mathbb{H}$ respectively. The linear tangent of an operator $\mathbb{A}$ is the directional derivative of the operator (the Gâteaux derivative):

$$\left( \frac{\partial \mathbb{A}}{\partial \widetilde{X}} \right) (dX) = \lim_{\beta \to 0} \frac{\mathbb{A}(\widetilde{X} + \beta dX) - \mathbb{A}(\widetilde{X})}{\beta}, \quad (8)$$

- $(\partial_X \mathbb{M})^*$ and $(\partial_X \mathbb{H})^*$ their *adjoint operators*. The adjoint $\mathbb{A}^*$ of a linear operator $\mathbb{A}$ on a space $\mathcal{D}$ is such as:

$$\forall x_1, x_2 \in \mathcal{D}, \ < \mathbb{A}x_1, x_2 > = < x_1, \mathbb{A}^* x_2 >, \quad (9)$$

it can be shown that canceling the gradient $\delta \mathcal{J}_X(\theta)$ with respect to the adjoint variables $\boldsymbol{\lambda}$ leads to a retrograde integration of an adjoint evolution model that takes into account the observations. Once the adjoint variables $\boldsymbol{\lambda}$ are estimated, one can recover the system state $X$ using relation (7). Finally, when dealing with non-linear models, recovering $X$ leads to the following incremental algorithm (Bennett, 1992):

---

1. Starting from $\tilde{X}(\boldsymbol{x}, t_0) = X_0(\boldsymbol{x})$, perform a *forward* integration: $\frac{\partial \tilde{X}}{\partial t} + \mathbb{M}(\tilde{X}) = 0$

2. $\tilde{X}(\boldsymbol{x}, t)$ being available, **compute the adjoint variables** $\boldsymbol{\lambda}(\boldsymbol{x}, t)$ with the *backward* equation:

   $\boldsymbol{\lambda}(t_f) = 0$ ;

   $-\frac{\partial \boldsymbol{\lambda}}{\partial t}(t) + (\partial_X \mathbb{M})^* \boldsymbol{\lambda}(t) = (\partial_X \mathbb{H})^* R^{-1} (\boldsymbol{Y} - \mathbb{H}(\tilde{X}))(t)$
   
   $\qquad\qquad\qquad\qquad\qquad\qquad\qquad (10)$

3. **Update the initial condition** : $dX(t_0) = B\boldsymbol{\lambda}(t_0)$;

4. $\boldsymbol{\lambda}$ being available, **compute the state space** $dX(t)$ from $dX(t_0)$ with the *forward* integration

   $$\frac{d dX}{dt}(t) + \left( \frac{\partial \mathbb{M}}{\partial \widetilde{X}} \right) dX(t) = Q\boldsymbol{\lambda}(t) \qquad (11)$$

5. **Update** : $\tilde{X} = \tilde{X} + dX$

6. **Loop** to step (2) until convergence

---

Intuitively, the adjoint variables $\boldsymbol{\lambda}$ contain information about the discrepancy between the observations and the dynamic model. They are computed from a current solution $\tilde{X}$ with the backward integration (10) that encompasses both the observations and the dynamic operators. This deviation indicator between the observations and the model is then used to

refine the initial condition (step (3)) and to recover the system state through an imperfect dynamic model where errors are $Q\lambda$ (step (4)). It should be noted that if the dynamic is perfect, the associated error covariance $Q$ is zero and the algorithm only refines the initial condition.

**Graphic Illustration of the Incremental Algorithm.** Let us try to illustrate in a simple way 'what is the algorithm doing'. For the sake of clarity, we will consider here the case of a prefect dynamic model where only the initial condition is refined.

-Let's take a closer look at (step (1)) of the above mentionned algorithm: it consists in determining a first coarse 'trajectory' of $X$. As shown in figure 2(a), we start from an initial condition $X(t_0)$ and compute the state variable values $\forall t \in [t_0, t_f]$ using a forward integration of the dynamical model $\mathbb{M}$. The red circles symbolically represent the available measurements $Y$.

-Having a first estimation of $X$, we may now move on to (step (2)), ie solving the backward integration that encompasses both the observations and the dynamic operators. As represented in figure 2(b), this step consists in integrating from $t_f$ to $t_0$ the adjoint dynamic model (10) and may be understood as a backpropagation of the discrepancy between our estimated trajectory and the observations $Y$.

-The purpose of computing the adjoint variable is to use its initial value $\lambda(t_0)$ and the relation defined in (step (3)), thus obtaining an initial increment $dX(t_0)$. We then aim at 'propagating' this latter along the time axis integrating forward the tangent model (11).

-Finally, updating the previous estimation with this increment should yield a better compromise between the dynamical model and the observations, figure 2(c)

**Remarks on Variationnal Assimilation.** An interresting property of variationnal assimilation is the mutual interraction of all estimations. That is for each $(t_1, t_2) \in [t_0, t_f]^2$ the estimated value of $X(t_1)$ at a given iteration can influence the value of $X(t_2)$ and vice versa. This yields to more coherent estimation than sequential assimilation where values of $X$ for a given time $t^*$ may only interract with values for $t > t^*$. This framework is then an appealing solution for large system states and acts as a smoothing issue. To design an assimilation process, we need to define:

1. The system state;

2. The dynamical model (and its adjoint);

3. The observation operator (and its adjoint);

4. The error covariance matrices.

   In this paper, we suggest to exploit this framework for defining an original multi-resolution system. The



(a) forward integration starting from inital condition (step (1))



(b) backward adjoint integration and initial correction (steps (2)-(3))



(c) forward tangent integration and correction propagation (steps (4)-(5))
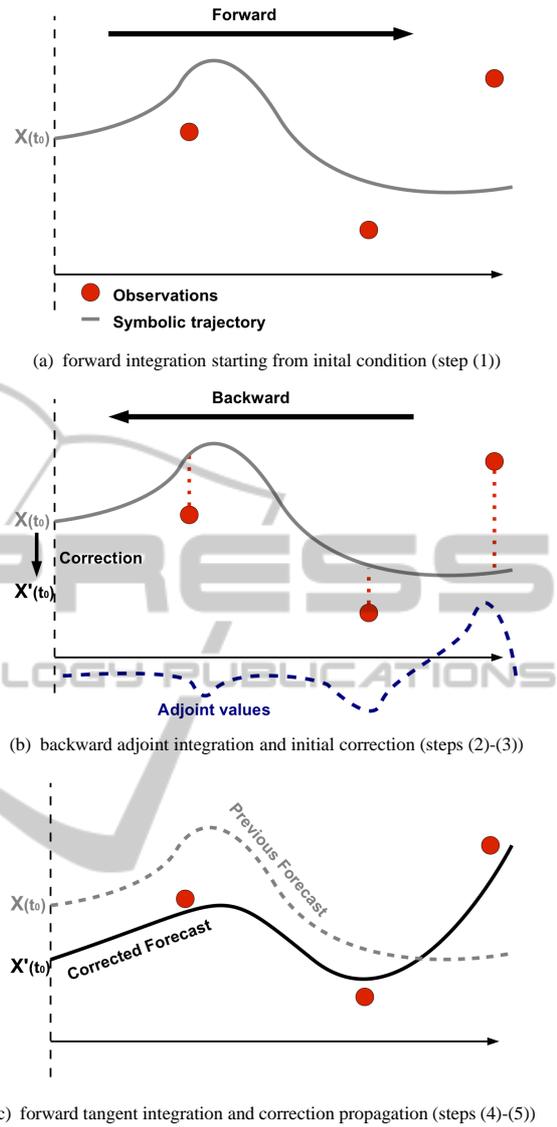
Figure 2: Symbolic illustration of one iteration of the incremental algorithm.

temporal variable will be related to the spatial scales, as presented in the next section.

# 3 VARIATIONAL ASSIMILATION FOR MULTI-RESOLUTION

Unlike most of existing issues related to variational assimilation, in this article the usual temporal variable $t$ is now connected to the different resolutions. To avoid ambiguities, we then rather prefer to represent its value with the scale parameter $\ell$. A value $\ell = 0$ corresponds to the "plain" resolution at the image grid

($\ell = 0 \Rightarrow$ grid $\boldsymbol{X}_0 = \boldsymbol{x}$) and we assume that the system state $X$ evolves across the resolutions following scale space relation:

$$\frac{\partial X}{\partial \ell} = \frac{1}{2}\Delta X. \qquad (12)$$

Indeed, it can easily be demonstrated that the solution of the previous relation is

$$X^\ell = X(\ell) = g_{\sqrt{\ell}} * X(0), \qquad (13)$$

where $g_{\sqrt{\ell}}$ is a gaussian kernel of standard deviation $\ell$. As shown in section 1, this relation enables to access the various scales of $X$. The linear model $\mathbb{M}(X) = -\frac{1}{2}\Delta X$ associated to relation (3) and corresponding to the scale space evolution (12) is then a perfect model for an exploration of $X$ at different resolution levels. We then suggest to rely on this relation, using the assimilation framework of the previous section, to define a multi-resolution system. Following the definition of adjoint operators, we get $(\partial_X \mathbb{M})^*(X) = \mathbb{M}(X) = -\frac{1}{2}\Delta X$.

The new multi-resolution procedure consists now to estimate $X(0)$ at the initial artificial time $\ell = 0$ (which corresponds to the image grid) under the perfect model of relation (12). The initial condition $X_0$ is set to zero and the observation system $\boldsymbol{Y}(\boldsymbol{X}^0) = \mathbb{H}(\boldsymbol{X}^0, X^0)$ defined at the image grid $\boldsymbol{X}^0$ reads for any resolution $\ell \geqslant 0$ :

$$\begin{cases} \boldsymbol{Y}(\boldsymbol{X}^\ell) = g_{\sqrt{\ell}} * \boldsymbol{Y}(\boldsymbol{X}^0) \\ \mathbb{H}(\boldsymbol{X}^\ell, X^\ell) = g_{\sqrt{\ell}} * \mathbb{H}(\boldsymbol{X}^0, X^0). \end{cases} \qquad (14)$$

Following the algorithm described in section 2, as the first integration of the dynamic model in (12) with null initial condition gives zeros for all $\ell$, the process is expressed as (with $B$ and $R$ the error covariance matrices related to the uncertainty on the initial condition and the observations):

---

- Initialization: $\tilde{X}(\ell) = 0$ for all scales $\ell \in [0, \ell_f]$.
1. Compute the adjoint variables $\boldsymbol{\lambda}(\ell)$ with the **downscaling** equation:

   $$\boldsymbol{\lambda}(\ell_f) = 0 \; ;$$
   $$-\frac{\partial \boldsymbol{\lambda}}{\partial \ell}(\ell) - \frac{1}{2}\Delta\boldsymbol{\lambda}(\ell) = \left(\partial_X \mathbb{H}^\ell\right)^* R^{-1} \qquad (15)$$
   $$(\boldsymbol{Y}(\boldsymbol{X}^\ell) - \mathbb{H}(\boldsymbol{X}^\ell, \tilde{X}^\ell))$$

2. Adjoint variables $\boldsymbol{\lambda}$ being available, update the **correction at the image grid** : $dX(0) = B\boldsymbol{\lambda}(0)$;
3. Assess to the correction at all the resolution levels $dX(\ell)$ from $dX(0)$ with the **upscaling** equation:

   $$\frac{\partial dX}{\partial \ell} = \frac{1}{2}\Delta dX \qquad (16)$$

4. **Update** : $\tilde{X} = \tilde{X} + dX$, $\forall \ell \in [0, \ell_f]$
5. **Loop** to step (1) until convergence

---

**Comments of the new multi-resolution scheme**: the first step of the previous algorithm corresponds to the usual downscaling approach: from coarse to fine resolutions, we compute and propagate the errors with equation (15). This enables to refine the solution at the image grid (step 2). Unlike the classic multi-resolution approach which ends here, we then re-propagate this correction at the various resolutions (step 3). The process is then repeated until convergence. This framework enables to modify former solutions at given resolution $\ell = L$ by taking into account their influence on smaller scales $\ell < L$. This answers a difficulty mentioned in section 1 and authorizes a correction of the estimations at all the resolution levels. As the dynamical model is perfect, for each application we need to define an observation system $\boldsymbol{Y}(\boldsymbol{X}^\ell) = \mathbb{H}(\boldsymbol{X}^\ell, X^\ell)$, its associated error covariance matrix $R$ and the one related to the null initial condition $B$. In the next section, we apply this framework for optical-flow estimation.

# 4 APPLICATION TO OPTICAL-FLOW

The goal of this section is to demonstrate the efficiency of the introduced multi-resolution procedure. To that end, we use the context of optical flow estimation and we compare results reached using two classic and our new multi-resolution approaches.

The optical-flow problem consists in estimating a velocity field $X(\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x}) = [u(\boldsymbol{x}), v(\boldsymbol{x})]^T$ at the image grid $\boldsymbol{x}$ between two images $I_1(\boldsymbol{x})$ and $I_2(\boldsymbol{x})$.

As mentioned above, the framework of section 3 is valid for any kind of applications, depending on the observation operator $\mathbb{H}$. In this example related to motion estimation, we use a simple Lucas-Kanade, presented in the next paragraph. It should be outlined that this observation model is implemented for a testing issue but obviously, one of the main advantage of this framework is its possibility to naturally embed in the observation term more advances techniques, as continuity preserving ones or some physical models accurately defined at a given resolution $\ell$ (sub-grid models in particular).

## 4.1 Observation Operator based on Lucas-Kanade

The most used and simple observation model proposed for optical-flow estimation is the brightness consistency assumption:

$$\frac{dI}{dt} = \frac{\partial I(\boldsymbol{x}, t)}{\partial t} + \boldsymbol{v}(\boldsymbol{x}, t) \cdot \nabla I(\boldsymbol{x}, t) \sim 0 \qquad (17)$$

and assumes that the points $\boldsymbol{x}$ keep their intensity along their displacements, the luminance $I$ being viewed as a continuous function and $\nabla = (\partial/\partial x, \partial/\partial y)^T$ being the gradient operator. Applied to a pair of images this relation reads:

$$
\begin{aligned}
I_2(\boldsymbol{x} + \Delta t \boldsymbol{v}(\boldsymbol{x})) - I_1(\boldsymbol{x}) = 0 &\Rightarrow \\
\frac{I_2(\boldsymbol{x}) - I_1(\boldsymbol{x})}{\Delta t} + \boldsymbol{v}(\boldsymbol{x}) \cdot \nabla I_2(\boldsymbol{x}) &= 0
\end{aligned}
\quad (18)
$$

where we have a first order Taylor development of the conservation constraint $I_2(\boldsymbol{x} + \Delta t \boldsymbol{v}(\boldsymbol{x})) - I_1(\boldsymbol{x}) = 0$ around $\Delta t \boldsymbol{v}(\boldsymbol{x})$ and $\Delta t$ is the time between two images (by convention we assume $\Delta t = 1$). This creates a link between the displaced frame difference $I_2(\boldsymbol{x}) - I_1(\boldsymbol{x})$, the spatial gradients of the second image $\nabla I_2(\boldsymbol{x})$ and the velocity $\boldsymbol{v}$ to estimate. The equations in (17-18) are commonly named the *optical-flow constraint equations* (OFCE) and are the basis of huge amount of studies. The reader can refer to (Baker and Matthews, 2004; Baker et al., 2007; Baker et al., 2010; Barron et al., 1994; Galvin et al., 1998; Mitiche and Bouthemy, 1996) for presentations and overviews of optical flow techniques. At this step, it is easy to observe that:

1. in homogeneous areas, all terms vanish and there is an infinity of solutions;

2. because of the projection $\boldsymbol{v} \cdot \nabla I$, only the normal component to the photometric gradients can be extracted with such a formulation. This problem is known as the "aperture problem".

Therefore, the relations in (17-18) are in themselves not sufficient to extract the velocity field. We need to add some constraints on the velocity to estimate.

In the work of Lucas & Kanade (Lucas and Kanade, 1981), the authors have assumed for each location $\boldsymbol{x}$ that the velocity is locally constant. It is then estimated as:

$$
\boldsymbol{v} = \min_{\boldsymbol{v} = (u,v)^T} \int_\Omega g_\sigma * \left( \frac{\partial I(\boldsymbol{x},t)}{\partial t} + \boldsymbol{v}(\boldsymbol{x},t) \cdot \nabla I(\boldsymbol{x},t) \right)^2 d\boldsymbol{x},
\quad (19)
$$

where $g_\sigma$ is a Gaussian window of standard deviation $\sigma$ in which the velocity $\boldsymbol{v}$ is assumed to be homogeneous. Canceling the derivative of the previous relation with respect to $\boldsymbol{v}$ leads to:

$$
\boldsymbol{v} = - \left( g_\sigma * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right)^{-1} g_\sigma * \begin{bmatrix} I_x I_t \\ I_y I_t \end{bmatrix}, \quad (20)
$$

where $I_\bullet = \partial I/\partial \bullet$. To guarantee a good conditioning of the previous matrix to invert, the spatial gradients must not vanish. The gaussian smoothing aims in fact at alleviating homogeneous areas by capturing the spatial information at a scale related to $\sigma$. On the

basis of relation (20), we then define our observation system where the unknown $X^\ell$ can be set at each resolution $\boldsymbol{X}^\ell$ with the linear Lucas-Kanade based observation system $\boldsymbol{Y}(\boldsymbol{X}^\ell) = \mathbb{H}(\boldsymbol{X}^\ell) X^\ell$ as:

$$
\begin{cases}
\boldsymbol{Y}(\boldsymbol{X}^\ell) = g_{\sqrt{\ell}} * g_\sigma * \begin{bmatrix} I_x I_t \\ I_y I_t \end{bmatrix} \\
\mathbb{H}(\boldsymbol{X}^\ell) = -g_{\sqrt{\ell}} * \left( g_\sigma * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right).
\end{cases}
\quad (21)
$$

The adjoint operator of $\mathbb{H}$ is itself and $g_\sigma$ is a convolution with a gaussian of standard deviation $\sigma$ related to the Lucas & Kanade strategy.

## 4.2 Error Covariance Matrix and Initialization Issues

**Error covariance matrix related to the observation**: we have defined, at each point of the grids $\boldsymbol{X}^\ell$, the matrix $R^{-1}$ used in (10) as a diagonal one such that:

$$
R^{-1}(\boldsymbol{X}^\ell) = R_{max} \exp - \frac{[\boldsymbol{Y}(\boldsymbol{X}^\ell) - \mathbb{H}(\boldsymbol{X}^\ell)]^2}{\sigma_{obs}^2}
\quad (22)
$$

As shown in (Corpetti et al., 2009), this penalization amounts to consider a robust norm in relation (19). Such a robust function allows the discarding of points having large "residual" values of the observation error $[\boldsymbol{Y}(\boldsymbol{X}^\ell) - \mathbb{H}(\boldsymbol{X}^\ell)]$ (called *outliers* in the Robust Statistics literature (Huber, 1981; Geman and Reynolds, 1992; Delanay and Bresler, 1998)). In our application, it enables to properly deal with corrupted areas that do not fit our data model exactly.

**Initial conditions:** as mentioned in section 3, we have no prior knowledge $X_0$. We then have set $X_0 = 0$. Therefore the associated covariance matrix $B$ is

$$
B(\boldsymbol{X}^0) = Id - \exp\left\{ -|I_2(\boldsymbol{X}^0) - I_1(\boldsymbol{X}^0)|^2 / \sigma_B^2 \right\} \quad (23)
$$

where $\sigma_B$ is a parameter to fix. It is related to the adequacy of the null solution depending on the OFCE.

**Complete scheme:** the multi-resolution strategy of section 3 with the observation operator defined in (21) and the associated error covariance matrixes $R$ and $B$ (defined in (22) and (23) respectively) constitute the complete framework for optical-flow estimation using an original downscaling/upscaling multi-resolution process. Let us now turn to some experimental results.

Table 1: **Quantitative comparisons on the DNS sequence** with a Pyramidal Lucas-Kanade (LK, (Lucas and Kanade, 1981)), a Lucas-Kanade term embed in a multi-resolution scheme with successive convolutions without any assimilation (CONV) and the proposed multi-resolution using variational assimilation (AMR). In addition, we have plotted numerical values issued from a commercial technique based on correlation (COM, LA VISION SYSTEM), Horn & Schunck (HS, (Horn and Schunck, 1981)), two fluid dedicated motion estimators with div-curl smoothing terms (DC 1 : (Corpetti et al., 2002); DC2 : (Yuan et al., 2007)), a stochastic observation term (STO, (Corpetti and Mémin, 2012)).

| | PYR | CONV | **AMR** | COM | HS | DC 1 | DC 2 | STO |
|---|---|---|---|---|---|---|---|---|
| AAE | $6.07^o$ | $4.53^o$ | $\mathbf{3.74^o}$ | $4.58^o$ | $4.27^o$ | $4.35^o$ | $3.04^o$ | $3.12^o$ |
| RMSE | 0.1699 | 0.1243 | **0.1057** | 0.1520 | 0.1385 | 0.1340 | 0.09602 | 0.0961 |

## 5 EXPERIMENTAL RESULTS

We have tested the technique presented in this paper on three kind of synthetic images:

1. **Synthetic fluid particles:** it corresponds to a synthetic pair of images issued from Direct Numerical Simulation of Navier-Stokes equations. More precisely the sequence simulates the evolution of particles submitted to a 2D turbulent flow. We have used such data since it is known that turbulent flows exhibit many interactions between the different scales. Therefore it is expected that the benefit of the approach will be highlighted.

2. **Synthetic natural image:** it corresponds to a synthetic pair of images generated by hand with simple and discontinuous motions. On this images, some areas are submitted to the aperture problem and we will show that the introduced multi-resolution is able to improve the results.

3. **Synthetic image issued from the middleburry database:** the Middleburry database has recently been proposed in (Baker et al., 2007) to compare recent and state-of-the-art optical-flow methods. It contains several sequences with various challenging situations like hidden textures, complex scenes, non rigid motion, high motion discontinuities, ... We have tested the technique on one pair issued from this dataset.

In the experiments, the results obtained by our technique are noted AMR for "Assimilation Multi-Resolution". We recall here that the objective of this section is more to validate the multi-resolution strategy rather than proposing an efficient optical-flow estimation technique. Therefore we have compared our AMR technique with two other Lucas-Kanade estimators but using different multi-resolution approaches: PYR for the usual pyramidal decomposition and CONV for a downscaling approach on successive convolutions (see the introduction).

### 5.1 Synthetic Fluid Particles

In the top of figure 3, we present: one image of the sequence (fig. 3(a)), the ground truth (fig. 3(b)), an estimated velocity field with the CONV technique (fig. 3(c)) and with the AMR one (fig. 3(d)). As one can observe on the velocity fields, all are similar and closed to the ground truth. However when observing some quantitative values of AAE (Average Angular Error) and RMSE (Root Mean Square Error) in table 1, it is promising to observe that among the three techniques PYR, CONV and AMR that correspond to the same estimator with different multi-resolution strategies, the proposed one is the most performing. It is also interesting to note that the pyramidal technique usually exploited to obtain the various scales is less performing than a series of convolutions. In addition, we have reported on this table some published results, on this pair of images, of other techniques especially devoted to such fluid and particle flows. It can be pointed out from these values that compared to more advanced approaches (related to optical-flow or devoted to fluid images), the improvement obtained with this framework yields this technique very competitive, despite a very simple observation term. We can then conclude from this experience that most of the common multi-resolution techniques are likely to introduce some errors that can partially be removed with a more advanced strategy, as the one presented in this paper.

### 5.2 Synthetic Natural Image

In the middle of figure 3, we present: one image of the sequence (fig. 3(e)), the ground truth (fig. 3(f)), an estimated velocity field with the CONV technique (fig. 3(g)) and with the AMR one (fig. 3(h)). As shown in fig. 3(f), the synthetic velocity field is composed two two distinct homogeneous motions (applied to the areas in dotted red lines in fig. 3(e)). One can observe that in the right hand part of the image, a constant region is submitted to the aperture problem and yields therefore the estimation very delicate. This difficulty
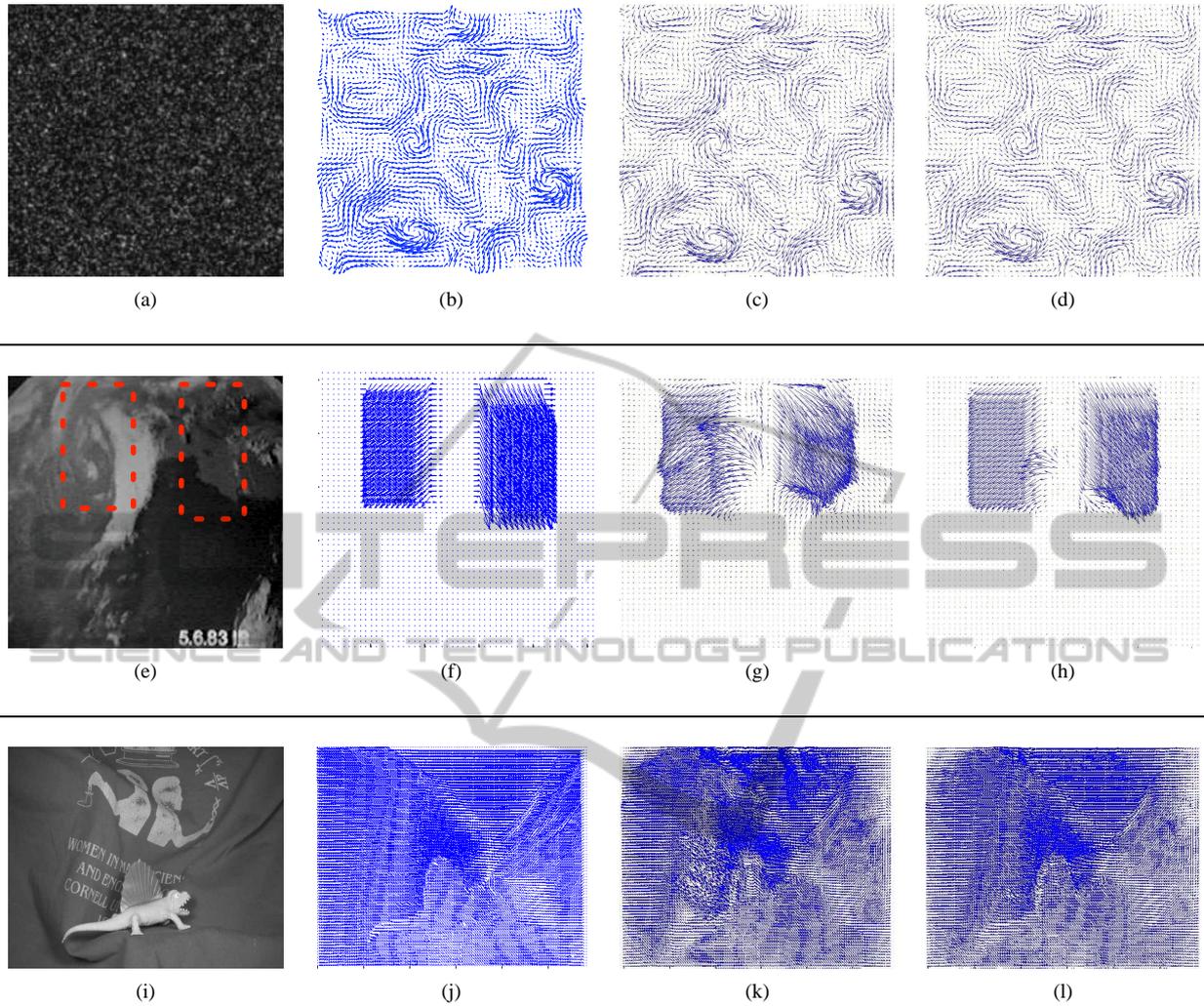
Figure 3: **Experiments on synthetic data**. Top: synthetic DNS, Middle: synthetic general image made by hand; Bottom: dimetrodon issued from the Middleburry database. For each pair, we respectively show one image of the sequence (a-e-i), the synthetic velocity field (b-f-j); one estimation using the CONV technique (c-g-k) and one estimation using the proposed AMR multi-resolution (d-h-l)

clearly affects the quality of the motion fields of figures 3(g–h). However from these two images, it appears clearly that the one issued from the AMR approach is better, as confirmed by the quantitative values of table 2. On this example, the first estimation at the coarser level had some errors. There latter are not corrected with a classical downscaling approach whereas they are modified using our technique.

## 5.3 Synthetic Images of the Middleburry Database

We have tested our approach on the "Dimetrodon" pair of image issued from the Middleburry database. The figure 3 embeds one image of the sequence (fig.

Table 2: **Quantitative comparisons on the synthetic sequence** with a Lucas-Kanade term embed in a multi-resolution scheme with successive convolutions without any assimilation (CONV) and the proposed multi-resolution using variational assimilation (AMR).

|        | CONV      | **AMR**             |
|--------|-----------|---------------------|
| AAE    | $9.02^o$  | **$5.62^o$**        |
| RMSE   | 1.28      | **0.79**            |

3(i)), the ground truth (fig. 3(j)), an estimated velocity field with the CONV technique (fig. 3(k)) and with the AMR one (fig. 3(l)). Associated numerical values are depicted in table 3. Here again, on this complex motion with many discontinuities, the original multi-resolution introduced in this paper outperforms

Table 3: **Quantitative comparisons on the dimetrodon sequence** with a Lucas-Kanade term embed in a multi-resolution scheme with successive convolutions without any assimilation (CONV) and the proposed multi-resolution using variational assimilation (AMR).

|      | CONV    | **AMR** |
|------|---------|---------|
| AAE  | $7.95^o$ | $\mathbf{6.50^o}$ |
| RMSE | 0.98    | **0.62** |

a usual one using similar observation terms. This is a promising behavior.

## 5.4 Key Message

Various situations have been tested with these three experiments: the first one exhibits a flow with many interactions between scales, the second one is submitted to the aperture problem whereas the third one is composed of a more complex velocity field. From the associated qualitative and quantitative values, it is very interesting to point out that for a similar technique (i.e. the Lucas & Kanade estimaion), the new multi-resolution approach is more competitive in all applications. This is the key message of these experiments.

## 6 CONCLUSIONS

In this paper, we have introduced an original mean to perform multi-resolution strategies commonly used in computer vision. These techniques, used to manage efficiently some simplifications (as linearization), generally suffer from one drawback: their inability to correct errors of coarser resolutions. Errors are indeed most of the time propagated along the scales. In this study, we have exploited a framework issued from optimal control theory and in particular variational data assimilation to solve this issue. The general idea of variational data assimilation techniques consist in performing a set of forward/backward integrations of a dynamical system to estimate a system state. Applied to a scale-space equations, we have derived a consistent mathematical framework to perform any multi-resolution scheme in a set of forward/backward integrations that in practice correspond to a set of downscaling/upscaling estimations.

We have validated the idea on a simple Lucas-Kanade motion estimation technique for three synthetic pair of images corresponding to various situations. The experimental results reveal that for all tested images, our multi-resolution approach outperforms classic ones, which is a very interesting and promising conclusion. As future works, we will

use more advanced observation terms associated with non-linear scale space dynamics able to preserve discontinuities.

## REFERENCES

Alvarez, L., Weickert, J., and Sánchez, J. (2000). Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1):41–56.

Baatz, M. and Schaape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In Strobl, J., editor, *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000, Karlsruhe, Herbert Wichmann Verlag*, pages 12–23.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1 – 21.

Baker, S. and Matthews, I. (2004). Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., and Szeliski, R. (2007). A Database and Evaluation Methodology for Optical Flow. In *Int. Conf. on Comp. Vis., ICCV 2007*.

Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., and Szeliski, R. (2010). A Database and Evaluation Methodology for Optical Flow. *International Journal of Computer Vision*, 92(1):1–31.

Barron, J., Fleet, D., Beauchemin, S., and Burkitt, T. (1994). Performance Of Optical Flow Techniques. *International Journal of Computer Vision*, 12(1):43–77.

Bennett, A. (1992). *Inverse Methods in Physical Oceanography*. Cambridge University Press.

Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. pages 25–36. Springer.

Burt, P. (1988). Smart sensing within a pyramid vision machine. *Proceedings of the IEEE*, 76(8):1006 – 1015.

Corpetti, T., Héas, P., Mémin, E., and Papadakis, N. (2009). Pressure image assimilation for atmospheric motion estimation. *Tellus Series A: Dynamic Meteorology and Oceanography*, 61(1):160–178.

Corpetti, T. and Mémin, E. (2012). Stochastic uncertainty models for the luminance consistency assumption. *IEEE Trans. on Image Processing, to appear*.

Corpetti, T., Mémin, E., and Pérez, P. (2002). Dense Estimation of Fluid Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):365–380.

Delanay, A. and Bresler, Y. (1998). Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. *IEEE Trans. Image Processing*, 7(2):204–221.

Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., and Stuetzle, W. (1995). Multiresolution analysis of arbitrary meshes. In *Proceedings of the 22nd*

*annual conference on Computer graphics and inter-active techniques*, SIGGRAPH '95, pages 173–182, New York, NY, USA. ACM.

Galvin, B., McCane, B., Novins, K., Mason, D., and Mills, S. (1998). Recovering motion fields: an analysis of eight optical flow algorithms. In *Proc. British Mach. Vis. Conf.*, Southampton.

Geman, D. and Reynolds, G. (1992). Constrained Restoration and The Recovery Of Discontinuities. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(3):367–383.

Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203.

Huber, P. (1981). *Robust Statistics*. John Wiley & Sons.

Le Dimet, F.-X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, pages 97–110.

Lions, J.-L. (1971). *Optimal control of systems governed by PDEs*. Springer-Verlag.

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *International joint conference on artificial*, 130:121–130.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693.

Mitiche, A. and Bouthemy, P. (1996). Computation of image motion: a synopsis of current problems and methods. *Int. Journ. of Comp. Vis.*, 19(1):29–55.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:971–987.

Talagrand, O. (1997). Assimilation of observations, an introduction. *J. Meteor. Soc. Jap.*, 75:191–209.

Talagrand, O. and Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. {I}: Theory. *J. of Roy. Meteo. soc.*, 113:1311–1328.

Vidard, P., Blayo, E., Le Dimet, F.-X., and Piacentini, A. (2000). 4D Variational Data Analysis with Imperfect Model. *Flow, Turbulence and Combustion*, 65(3-4):489–504.

Yuan, J., Schnoerr, C., and Mémin, E. (2007). Discrete orthogonal decomposition and variational fluid flow estimation. *Journ. of Mathematical Imaging and Vision*, 28(1):67–80.