A proper generalized decomposition approach for optical flow estimation

Abdallah El Hamidi, Marwan Saleh, Nicolas Papadakis, Baudouin Denis de
Senneville

A PROPER GENERALIZED DECOMPOSITION APPROACH FOR OPTICAL FLOW ESTIMATION

A. EL HAMIDI[*†], M. SALEH[†], N. PAPADAKIS[‡], AND B. DENIS DE SENNEVILLE[*‡]

Abstract. This paper introduces the use of the Proper Generalized Decomposition (PGD) method for the optical flow (OF) problem in a classical framework of Sobolev spaces, *i.e.* optical flow methods including a robust energy for the data fidelity term together with a quadratic penalizer for the regularisation term.

A mathematical study of PGD methods is first presented for general regularization problems in the framework of (Hilbert) Sobolev spaces, and their convergence is then illustrated on OF computation. The convergence study is divided in two parts: (i) the weak convergence based on the Brézis-Lieb decomposition, (ii) the strong convergence based on a growth result on the sequence of descent directions. A practical PGD-based OF implementation is then proposed and evaluated on freely available OF data sets.

The proposed PGD-based OF approach outperforms the corresponding non-PGD implementation in terms of both accuracy and computation time for images containing a weak level of information, namely low image resolution and/or low Signal-To-Noise Ratio (SNR).

Key words. Proper Generalized Decomposition (PGD) , AND Optical Flow , AND Optimization

AMS subject classifications.

1. Introduction. Separated representations were introduced in the $80's$ by P. Ladevèze who proposed a space-time separated representation of transient solutions involved in strongly non-linear models [18]. We refer the interested reader to Ladeveze and co-authors works on this subject [20, 19, 21, 11]. Later, separated representations were employed in the context of stochastic modeling [23] as well as for solving multidimensional models suffering from the so-called "curse of dimensionality", some of them never solved before [2, 9]. The techniques making use of separated representations computed on the fly were called Proper Generalized Decompositions (PGD) for defining very efficient algorithms able to minimize smooth and convex functionals.

In this work, we are interested on the variational optical flow (OF) application, initially proposed in the context of motion estimation in video sequences by Horn&Schunck in 1981 [17]. The latter assumes that pixels conserve their intensity along their trajectory, to which a spatial regularity constraint of the estimated motion is added [15]. Classical methods then consider the estimation of the optical flow, a $2D$ vector field mapping the pixel of one $2D$ image to another one, through the minimization of an energy functional containing two quadratic (convex and smooth) terms, penalizing the intensity conservation constraint and the motion regularity. Using robust (*i.e.* non-quadratic) penalizers applied on the data fidelity term, it is expected that the estimated motion will be less prone to errors caused by local gray-level intensity perturbations. Such a model was recently proved to be of great interest for the OF-based estimation of organ displacement during MR-guided surgical operations [30]. Current state-of-the art methods in computer vision also use robust regulariz-

[†]"Laboratoire des Sciences de l'Ingénieur pour l'Environnement" (LaSIE), University of La Rochelle, France (aelhamid@univ-lr.fr, msaleh01@hotmail.com).

[‡]"Institut de Mathématiques de Bordeaux" (IMB), UMR 5251 CNRS/University of Bordeaux, Talence, France (npapadak@math.u-bordeaux.fr, bdenisde@math.u-bordeaux.fr).

ers (*e.g.* non-smooth Total Variation) [8, 29, 26], which are good priors when the motion is assumed to be piecewise constant. Such regularizers are therefore not always adapted for deformable objects and smoother regularizers may be preferable for particular applications such as organ registration in medical imaging [16].

Hence, this paper introduces the use of the Proper Generalized Decomposition (PGD) method for fast and accurate resolution of the Optical flow problem. We focus our interest in a classical framework of the (vectorial) Sobolev space $\left[H^1(\Omega)\right]^d$, where Ω is a bounded domain in $\mathbb{R}^d$, d being the dimension of Ω ($d = 2$ in the case of OF). For this purpose, we retrain our study to a global energy as a sum of a data term (with L^2 norm or smoothed L^1 norm) and a regularization term (with L^2 norm of the OF gradient). In this manuscript, we show how to use the tensorial character of the space $H^1(\Omega)$ in order to reduce significantly the number of unknowns from the total number of pixels $\prod_{k=1}^{d} N_k$ to $\sum_{k=1}^{d} N_k$, when $\Omega = \Omega_1 \times \cdots \Omega_d$ and N_k is the number of discretized points of the $1D$ interval Ω_k. We eventually show that such process can even improve the properties of flow diffusion by better taking into account interactions between neighboring pixels.

In the first part of the manuscript, the theoretical convergence of the method is demonstrated and a numerical implementation is proposed. The proposed PGD-based OF approach is then compared to the corresponding non-PGD implementation in terms of both accuracy and computation time on the freely available Middlebury image collection [4]. Benefits and drawbacks of the proposed PGD-based approach are then compared and discussed for various conditions of image resolution, SNR and number of color channels. It will be demonstrated that very fast and accurate OF algorithms can be designed for low resolution images using tensorial representation. As we will show, accurate motion fields can be estimated using a few linear combinations of tensor products.

Such works may be of major interest for real-time image-based motion estimation on embedded systems, especially when the latter require acute estimates with high framerate and short latency. In particular, it may meet the need of using cheap chipset for economical considerations, in order to expand the distribution of a product. Another practical example can be found in the specific context of MR-guided surgical operations which require reliable organ motion estimates using low resolution noisy images with high framerate and short latency [13].

2. Mathematical tools and methods.

2.1. Proposed PGD Approach for variational problems.
This section deals with minimization problems with convex functionals defined on tensor Banach spaces. Tensorial structure of Banach spaces allows optimization via dimensionality reduction methods in a progressive way. In what follows, we focus on PGD methods introduced by Ladevèze [18] and Ammar in [2, 3]. More precisely, we focus on the mathematical and numerical study of such methods in the variational framework.

2.1.1. Tensor Banach Spaces and PGD.
Consider a finite family of real Banach spaces $(V_k , \| \cdot \|_k)_{1 \leq k \leq d}$, where $d \geq 2$ is an integer. The algebraic tensor Banach space spanned by the family $(V_k , \| \cdot \|_k)_{1 \leq k \leq d}$, denoted by

$$V := {}_a\bigotimes_{k=1}^{d} V_k,$$

is the set of all elementary tensors finite linear combinations $v = \bigotimes_{k=1}^{d} v_k$, with $v_k \in V_k$.

The suffix "a" in "$_a\bigotimes$" refers to the "algebraic" nature of the tensor product. That is,

(2.1)
$$V = \left\{ \sum_{j=1}^{m} v_1^{(j)} \otimes v_2^{(j)} \otimes \cdots \otimes v_d^{(j)} \quad : \quad m \in \mathbb{N}^* \text{ and } v_k^{(j)} \in V_k \text{ for } k = 1, 2, \cdots, d \right\}.$$

We say that $V_{\|\cdot\|}$ is a Banach tensor space if there exists an algebraic tensor space V and a norm $\|\cdot\|$ on V such that $V_{\|\cdot\|}$ is the completion of V with respect to the norm $\|\cdot\|$, *i.e.*

$$V_{\|\cdot\|} = {}_a\bigotimes_{k=1}^{d} V_k \overline{\phantom{\bigotimes_{k=1}^{d} V_k}}^{\,\|\cdot\|} \quad .$$

In the special case where $V_{\|\cdot\|}$ is a Hilbert space, one says that $V_{\|\cdot\|}$ is a tensor Hilbert space.

In the sequel, the cone of all elementary tensors in V will be denoted by $\mathcal{R}_1(V)$, *i.e.*

(2.2) $\qquad \mathcal{R}_1(V) = \{v_1 \otimes v_2 \otimes \cdots \otimes v_d \quad : \quad v_k \in V_k \text{ for } k = 1, 2, \cdots, d\}.$

Thus the space spanned by $\mathcal{R}_1(V)$ is V which is in turn a dense subset of $V_{\|\cdot\|}$.

Now, consider a functional $J : V_{\|\cdot\|} \longrightarrow \mathbb{R}$, where $V_{\|\cdot\|}$ is a reflexive tensor Hilbert space. We are interested in the minimization problem

(2.3)
$$\inf_{u \in V_{\|\cdot\|}} J(u),$$

where J satisfies the hypotheses

(H_1) J is differentiable,
(H_2) J is coercive and $\widetilde{J} := J - J(0) - J'(0)$ is 2-elliptic, that is,

(2.4) $\exists\, C > 0 \; : \; \forall\, u, v \in V_{\|\cdot\|} \,$, one has $\langle \widetilde{J}'(u) - \widetilde{J}'(v) \,, \, u - v \rangle_{V_{\|\cdot\|}} \geq C\|u - v\|^2$,

where $\langle \cdot, \cdot \rangle_{V_{\|\cdot\|}}$ stands for the inner product in the Hilbert space $V_{\|\cdot\|}$. Notice that the monotony condition (2.4) on $\widetilde{J}'$ implies the strict convexity of J, but J is not 2-elliptic as in [14].

Applying standard arguments of the calculus of variations, one shows that problem (2.3) admits a unique solution, under hypotheses $(H_1) - (H_2)$. To define PGD methods, we assume furthermore

$$(H_3) \quad \mathcal{R}_1(V) \text{ is weakly closed in } V_{\|\cdot\|},$$

where $V := {}_a\bigotimes_{k=1}^{d} V_k$ is the algebraic tensor space associated to $V_{\|\cdot\|}$.

Notice that for any $u \in V$, the minimization problem

$$(2.5) \qquad\qquad \inf_{z \in \mathcal{R}_1(V)} J(u + z)$$

has at least one solution, under hypotheses $(H_1) - (H_3)$. The possible non-uniqueness of the solution to problem (2.5) is due to the fact that $\mathcal{R}_1(V)$ is not convex.

The PGD methods associated to problem (2.3) consist in the construction of a sequence $(u_m)_{m \in \mathbb{N}} \subset V$ as follows

    (i) Initialization: $u_0 := 0$.

    (ii) Descent direction: choose $\widehat{z}_m := \operatorname{argmin}_{z \in \mathcal{R}_1(V)} J(u_{m-1} + z)$.

    (iii) Update strategies

        <u>Strategy (c)</u>
        $u_m := u_{m-1} + \widehat{z}_m$.

        <u>Strategy (r)</u>
        $\star$ Construct a closed subspace of $V_{\|\cdot\|}$ containing $\widehat{z}_m$, denoted by $U(\widehat{z}_m)$.
        $\star$ Choose the descent direction: $\widetilde{z}_m := \operatorname{argmin}_{z \in U(\widehat{z}_m)} J(u_{m-1} + z)$.
        $\star$ Update $u_m := u_{m-1} + \widetilde{z}_m$.

        <u>Strategy (ℓ)</u>
        $\star$ Construct a closed subspace of $V_{\|\cdot\|}$ containing $u_{m-1} + \widehat{z}_m$, denoted by $U(u_{m-1} + \widehat{z}_m)$.
        $\star$ Update $u_m := \operatorname{argmin}_{u \in U(u_{m-1} + \widehat{z}_m)} J(u)$.

We refer the interested reader to [14] for possible choices of the closed subspaces $U(\widehat{z}_m)$ and $U(u_{m-1} + \widehat{z}_m)$. Thus, a PGD sequence $\mathbf{u} := (u_m)_{m \in \mathbb{N}}$ can be represented by a sequence, possibly finite, of symbols denoted by

$$\widehat{\alpha}(\mathbf{u}) = \alpha_1 \, \alpha_2 \, \alpha_3 \, \cdots \, \alpha_m \cdots$$

where $\alpha_m \in \{\ell, \, c, \, r\}$, for every $m = 1, \, 2, \, \cdots$

In what follows and without loss of generality, we will focus on a variational problem corresponding to the OF computation. Although we limit ourselves to the case of the OF, it is clear that our approach remains valid for any energy satisfying the hypotheses $(H_1) - (H_3)$ on a tensor Hilbert space.

2.1.2. Example of $L^2 - L^2$ Optical Flow. Let Ω be a domain of $\mathbb{R}^2$ which can be written as the Cartesian product of two open intervals $(\Omega_k)_{1 \leq k \leq 2}$ in $\mathbb{R}$.

Denote the "grey-level" function at the point $(x, y) \in \Omega$, at time t by $I(x, y, t)$ and

$$\Gamma(x_0, y_0, t_0) := \{(x(t), y(t)) \; : \; (x(t_0), y(t_0)) = (x_0, y_0), \; t \geq t_0\},$$

be the trajectory, supposed to be smooth, of the point (x, y). It is appropriate to assume that the grey level remains "constant" during an interval of infinitely small

length included in $\Gamma(x_0, y_0, t_0)$. This hypothesis can be written by

$$\frac{d\,I}{d\,t}(x(t), y(t), t) = 0, \quad \forall\, t > t_0,$$

and is commonly called "Optical Flow constraint". Using a Taylor approximation of order 1, this constraint can be rewritten as

$$\sigma(x(t), y(t), t) \cdot \nabla I\,(x(t), y(t), t) + \frac{\partial I}{\partial t}(x(t), y(t), t) = 0, \quad \forall\, t > t_0,$$

where $\sigma(x(t), y(t), t)$ is the flow of $(x(t), y(t))$ at time t and $\nabla I\,(x(t), y(t), t)$ is the spatial gradient of I at the point $(x(t), y(t))$. Hereafter, we set

$$\sigma := (u, v).$$

We fix a time $t > t_0$ and to simplify the notation the time dependence of the functions σ and I will be omitted.

We introduce the functional

(2.6)
$$J_0(\sigma) := \int_\Omega \left(\sigma(x, y) \cdot \nabla I\,(x, y) + \frac{\partial I}{\partial t}(x, y) \right)^2 \mathrm{d}x\,\mathrm{d}y.$$

It is classical to consider the following functional to estimate the OF

(2.7)
$$J(\sigma) := J_0(\sigma) + J_1(\sigma),$$

where

$$J_1(\sigma) := \int_\Omega c(x, y)|\sigma(x, y)|^2\,dx\,dy + \lambda \int_\Omega |\nabla \sigma(x, y)|^2\,\mathrm{d}x\,\mathrm{d}y.$$

The real parameter $\lambda > 0$ is a regularization parameter for $\nabla \sigma$ while the function $c(x, y) \geq \varepsilon > 0$, it penalizes homogeneous regions characterized by weak gradients in the image. Thus, $c(x, y)$ is small (resp. large) in regions of Ω where $|\nabla I(x, y)|$ is large (resp. small). We assume in what follows that $c \in \mathcal{C}^\infty(\overline{\Omega})$. This approach will be denoted by $L^2 - L^2$ approach, that is, the regularization term is taken with the $L^2(\Omega)-$norm and the same applies for the data-fidelity term. data

Remark *The term $\displaystyle\int_\Omega c(x, y)|\sigma(x, y)|^2\,dx\,dy$ in the regularization term $J_1(\sigma)$ is necessary for the (theoretical) coerciveness of the energy J, due to the lack of coerciveness in the regions of Ω where $\sigma(x, y) \cdot \nabla I\,(x, y) = 0$. In our applications, we observe that the absence of this term does not affect the quality the PGD convergence, that is, a "numerical coerciveness" of the energy holds true.*

It is obvious that the functional framework of J is $W_{\|\cdot\|} := \left[H^1(\Omega)\right]^2$ and consequently, the associated minimization problem for the OF estimate is

(2.8)
$$\inf_{\sigma \in W_{\|\cdot\|}} J(\sigma).$$

The usual norms on $W_{\|\cdot\|}$ and on $H^1(\Omega)$ will be denoted by $\|\cdot\|$.
Then we have

$$
(2.9) \qquad \|\xi\| = \begin{cases} \|\xi\|_2 + \|\nabla\xi\|_2 & \text{if} \quad \xi \in H^1(\Omega) \\[2ex] \| \,|\xi|^2\,\|_2^{1/2} + \| \,|\nabla\xi|^2\,\|_2^{1/2} & \text{if} \quad \xi \in W_{\|\cdot\|} \end{cases}
$$

where $\|\cdot\|_2$ denotes the usual norm on $L^2(\Omega)$, $|\cdot|$ is the Euclidean norm on $\mathbb{R}^2$ and $\nabla\xi := (\nabla\xi_1\,,\,\nabla\xi_2)$, for every $\xi = (\xi_1, \xi_2) \in W_{\|\cdot\|}$.

One can easily verify that the functional J satisfies the hypotheses $(H_1) - (H_2)$. Therefore, since the space $W_{\|\cdot\|}$ is reflexive, Problem (2.8) has a unique solution, by direct methods of variational calculus. Moreover, since the space $H^1(\Omega)$ has a Hilbert tensor structure, $i.e.,$

$$
(2.10) \qquad H^1(\Omega) = {}_a\overline{\bigotimes_{k=1}^{2} H^1(\Omega_k)}^{\,\|\cdot\|}\ ,
$$

then the space $W_{\|\cdot\|}$ also, as a finite cartesian product of tensorial Hilbert spaces. The algebraic tensor space associated to $W_{\|\cdot\|}$ is

$$
(2.11) \qquad W := \left[{}_a\bigotimes_{k=1}^{2} H^1(\Omega_k) \right]^2 .
$$

To describe the PGD method in the vectorial framework, we introduce the set

$$
(2.12) \qquad \mathcal{R}_1(W) := \left[\mathcal{R}_1\left({}_a\bigotimes_{k=1}^{2} H^1(\Omega_k) \right) \right]^2
$$

of elementary tensors in W.

It is well-known that $\mathcal{R}_1\left({}_a\bigotimes_{k=1}^{2} H^1(\Omega_k) \right)$ is weakly closed in $H^1(\Omega)$, consequently $\mathcal{R}_1(W)$ is also weakly closed in $W_{\|\cdot\|}$. Whence, the hypotheses $(H_1)-(H_3)$ are satisfied in the framework of OF considered here and for every $\sigma \in W_{\|\cdot\|}$, the minimization problem

$$
(2.13) \qquad \inf_{z\in\mathcal{R}_1(W)} J(\sigma + z)
$$

has a unique solution.

LEMMA 2.1. *If* $\sigma := (\sigma_1, \sigma_2) \in W_{\|\cdot\|}$ *satisfies*

$$
J(\sigma) \leq J(\sigma + z), \quad \forall z \in \mathcal{R}_1(W).
$$

then σ *is the unique solution of (2.8).*

Proof. Let $z = (\varphi, \psi) \in \mathcal{R}_1(W)$ and let the function $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by $g(r, s) = J(\sigma_1 + r\,\varphi, \sigma_2 + s)$. It is clear that the function g is differentiable on $\mathbb{R}^2$ and realizes is global minimum at $(0, 0)$. Consequently $\nabla g(0, 0) = J'(\sigma) \cdot (\varphi, \psi) = 0$. Using the linearity of $J'(\sigma)$, we deduce that $J'(\sigma) = 0$ on W. Finally the density of W in $W_{\|\cdot\|}$ concludes the proof. $\qquad \square$

We will consider the method PGD with the strategy (c) for the progressive approximation of the OF

$$(2.14) \quad (M) \quad \begin{cases} (i) \quad Initialization : \\ \quad\quad (u_0, v_0) := (0, 0) \ in \ W_{\|\cdot\|}. \\ \\ (ii) \quad Descent\,direction : \\ \quad\quad (\varphi_m, \psi_m) := \operatorname{argmin}_{(\varphi, \psi) \in \mathcal{R}_1(W)} J(u_{m-1} + \varphi, v_{m-1} + \psi). \\ \\ (iii) \quad Update : \\ \quad\quad (u_m, v_m) := (u_{m-1}, v_{m-1}) + (\varphi_m, \psi_m), \ \forall\, m \geq 1. \end{cases}$$

This is the most used PGD method thanks to its implementation simplicity. We prove in the sequel that every PGD sequence (M) is convergent in $W_{\|\cdot\|}$.

Remark The sequence of descent directions satisfy the identity

$$J'(u_m, v_m) \cdot (\varphi_m, \psi_m) = 0, \quad \forall\, m \geq 1.$$

THEOREM 2.2. *Let* $\Omega = \displaystyle\prod_{k=1}^{2} \Omega_k$ *be a domain in* $\mathbb{R}^3$*, where* $(\Omega_k)_{1 \leq k \leq 2}$ *denotes open intervals in* $\mathbb{R}$*, not necessarily bounded. Let* J *the functional defined on* $W_{\|\cdot\|}$ *by (2.7), where the function* $I \in W^{1,\infty}(]0, +\infty[\times\Omega)$ *and the function* c *is continuous on* $\overline{\Omega}$*, positive et bounded. Then the sequence* $(u_m, v_m)_{m \in \mathbb{N}} \subset W$*, defined by* (M)*, converges weakly in* $W_{\|\cdot\|}$ *toward the unique solution of (2.8).*

Proof. Let $(u_m, v_m)_{m \in \mathbb{N}} \subset W$ be a PGD sequence defined by (M). It is clear that the real sequence $(J(u_m, v_m))_{m \in \mathbb{N}}$ is decreasing, bounded below and consequently convergent in $\mathbb{R}$, we denote its limit by ℓ. The coerciveness of J implies that $(u_m, v_m)_{m \in \mathbb{N}}$ is bounded in $W_{\|\cdot\|}$.
To show that $(u_m, v_m)_m$ converges weakly in $W_{\|\cdot\|}$, we show that $(u_m, v_m)_m$ has a unique limit point for the weak topology of $W_{\|\cdot\|}$.
Let $(u, v) \in W_{\|\cdot\|}$ be a limit point of $(u_m, v_m)_m$ for the weak topology of $W_{\|\cdot\|}$ and let $(u_{m_k}, v_{m_k})_k$ a subsequence of $(u_m, v_m)_m$ converging weakly to (u, v).
As $u_{m_k} \rightharpoonup u$ in $H^1(\Omega_1)$ then $u_{m_k} + \varphi \rightharpoonup u + \varphi$, for every $\varphi \in H^1(\Omega_1)$. Therefore, we have the following Brézis-Lieb decompositions:

$$\|u_{m_k} - u\|^2_{L^2(\Omega_1)} = \|u_{m_k} + \varphi\|^2_{L^2(\Omega_1)} + \|u + \varphi\|^2_{L^2(\Omega_1)} - 2\,\langle u_{m_k} + \varphi,\, u + \varphi \rangle_{L^2(\Omega_1)}$$
$$= \|u_{m_k} + \varphi\|^2_{L^2(\Omega_1)} - \|u + \varphi\|^2_{L^2(\Omega_1)} + \mathrm{o}_k(1),$$

where $o_k(1)$ is a correction term satisfying $\displaystyle\lim_{k\to+\infty} o_k(1) = 0$, and

$$
\begin{aligned}
\|\nabla u_{m_k} - \nabla u\|^2_{L^2(\Omega_1)} &= \|\nabla u_{m_k} + \nabla\varphi\|^2_{L^2(\Omega_1)} + \|\nabla u + \nabla\varphi\|^2_{L^2(\Omega_1)} - \\
&\quad 2\langle \nabla\left(u_{m_k} + \varphi\right), \nabla\left(u + \varphi\right)\rangle_{L^2(\Omega_1)} \\
&= \|\nabla u_{m_k} + \nabla\varphi\|^2_{L^2(\Omega_1)} - \|\nabla u + \nabla\varphi\|^2_{L^2(\Omega_1)} + o_k(1).
\end{aligned}
$$
(2.15)

Using the same arguments for the sequences $(v_{m_k})_k$, we get

$$
J(u_{m_k} - u, v_{m_k} - v) = J(0) + J(u_{m_k}, v_{m_k}) - J(u, v) + o_k(1).
\tag{2.16}
$$

On the other hand,

$$
J(u_{m_{k+1}}, v_{m_{k+1}}) \leq J(u_{m_k+1}, v_{m_k+1}) \leq J(u_{m_k} + \varphi, v_{m_k} + \psi),
\tag{2.17}
$$
$$
\forall\, (\varphi, \psi) \in \mathcal{R}_1(W).
$$

Then, for every $(\varphi, \psi) \in \mathcal{R}_1(W)$ one gets

$$
J(u_{m_{k+1}} - u, v_{m_{k+1}} - v) + J(u, v) \leq J(u_{m_k} - u, v_{m_k} - v) + J(u+\varphi, v+\psi) + o_k(1).
\tag{2.18}
$$

Using the fact that $J(u_{m_{k+1}} - u, v_{m_{k+1}} - v)$ and $J(u_{m_k} - u, v_{m_k} - v)$ have the same limit $\ell - J(u, v)$ as $k \longrightarrow +\infty$, we obtain

$$
J(u, v) \leq J(u + \varphi, v + \psi), \quad \forall\, (\varphi, \psi) \in \mathcal{R}_1(W).
\tag{2.19}
$$

Lemma 2.1 implies that (u, v) is thus the unique solution (2.8). Hence, the sequence $(u_m, v_m)_{m\in\mathbb{N}}$ has a unique adherence value for the weak topology of $W_{\|\cdot\|}$, it converges weakly to the unique solution of problem (2.8). $\qquad\square$

In what follows, the dual space of $W_{\|\cdot\|}$ will be denoted by $W'_{\|\cdot\|}$ and the norm in this dual space by $\|\cdot\|_\star$.

COROLLARY 2.3. *Under the same hypotheses as in Theorem 2.2, one has*

$$
J'(\sigma_m) \rightharpoonup 0 \quad in\ W'_{\|\cdot\|} - weak\ \star.
$$

In other words,

$$
\lim_{m\to+\infty} J'(\sigma_m) \cdot z = 0, \quad \forall\, z \in W_{\|\cdot\|}.
$$

Proof. Simply write out the expression of $J'(\sigma_m) \cdot z$ and use the weak convergence of $(\sigma_m)_m$ to the unique solution of (2.8). $\qquad\square$

We give some technical lemmas in order to show the strong convergence of PGD sequences given by the scheme (M).

LEMMA 2.4. *The sequence of successive descent directions $z_m := (\varphi_m, \psi_m)$ defined in (M) for the PGD sequence $(\sigma_m)_m$ satisfies*

$$\sum_{k=1}^{+\infty} \|z_k\|^2 < +\infty,$$

in particular, $\lim\limits_{m \to +\infty} z_m = 0$ *in* $W_{\|\cdot\|}$.

Proof. Introducing the quadratic part of J defined on $W_{\|\cdot\|}$ by

$$\widetilde{J}(\sigma) := J(\sigma) - J(0) - J'(0) \cdot \sigma,$$

we have that

$$J'(\sigma) \cdot \chi = \widetilde{J}'(\sigma) \cdot \chi + J'(0) \cdot \chi.$$

By definition of $\widetilde{J}'(\sigma) \cdot \chi$ and due to the symmetry of the mapping $(\sigma, \chi) \mapsto \widetilde{J}'(\sigma) \cdot \chi$

$$
\begin{aligned}
J'(\sigma_m) \cdot \sigma_m &= J'(\sigma_m) \cdot \sigma_{m-1} \\
&= \widetilde{J}'(\sigma_m) \cdot \sigma_{m-1} + J'(0) \cdot \sigma_{m-1} \\
&= \widetilde{J}'(\sigma_{m-1}) \cdot \sigma_{m-1} + \\
&\quad J'(0) \cdot \sigma_{m-1} + \widetilde{J}'(z_m) \cdot \sigma_{m-1} \\
&= J'(\sigma_{m-1}) \cdot \sigma_{m-1} + \widetilde{J}'(\sigma_{m-1}) \cdot z_m \\
&\ \ \vdots \\
&= \sum_{j=1}^{m-1} \widetilde{J}'(\sigma_j) \cdot z_{j+1}, \quad \text{since } \sigma_0 = 0 \text{ and } \widetilde{J}'(0) = 0 \\
&= \sum_{1 \le i < j \le m} \widetilde{J}'(z_i) \cdot z_j, \quad \text{since } \sigma_i = \sum_{j=1}^{i} z_j \\
&= \frac{1}{2} \left(\widetilde{J}'(\sigma_m) \cdot \sigma_m - \sum_{k=1}^{m} \widetilde{J}'(z_k) \cdot z_k \right) \\
&= \frac{1}{2} \left(J'(\sigma_m) \cdot \sigma_m - J'(0) \cdot \sigma_m - \sum_{k=1}^{m} \widetilde{J}'(z_k) \cdot z_k \right).
\end{aligned}
$$

Hence, one obtains

$$(2.20) \qquad J'(\sigma_m) \cdot \sigma_m = -J'(0) \cdot \sigma_m - \sum_{k=1}^{m} \widetilde{J}'(z_k) \cdot z_k.$$

Since $\widetilde{J}'(z_k) \cdot z_k \ge C_1 \|z_k\|^2$, for any k and $(\sigma_m)_m$ bounded, we deduce by Corollary 2.3 that

$$(2.21) \qquad \sum_{k=1}^{m} \|z_k\|^2 \le C_2 \left(\|J'(\sigma_m)\|_\star \times \|\sigma_m\| + \|J'(0)\|_\star \times \|\sigma_m\| \right) \le C_3.$$

In particular, $\lim\limits_{m \to +\infty} \|z_m\| = 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

LEMMA 2.5. *Let $(\sigma_m)_{m \in \mathbb{N}}$ be a PGD sequence for the problem (2.8), there exists $C > 0$ such that*

$$|J'(\sigma_{m-1}) \cdot z| \leq C\|z_m\|\,\|z\|,$$

for any $z \in \mathcal{R}_1(W)$, where $z_m := (\varphi_m, \psi_m)$ is the descent direction at the iteration $m \geq 1$, defined in (M).

Proof. First, remark that $J' : W_{\|\cdot\|} \longrightarrow W^*_{\|\cdot\|}$ is Lipschitz continuous. On the other hand, the convexity of J leads to

$$(2.22)\qquad J'(\sigma_{m-1} + z) \cdot (z_m - z) \leq J(\sigma_{m-1} + z_m) - J(\sigma_{m-1} + z) \leq 0,$$

thus

$$(2.23)\qquad\qquad\qquad J'(\sigma_{m-1} + z) \cdot z_m \leq J'(\sigma_{m-1} + z) \cdot z.$$

$$
\begin{aligned}
-J'(\sigma_{m-1}) \cdot z &= (J'(\sigma_{m-1} + z) - J'(\sigma_{m-1})) \cdot z - J'(\sigma_{m-1} + z) \cdot z \\
&\leq C\|z\|^2 - J'(\sigma_{m-1} + z) \cdot z \quad (J \text{ is Lipschitz continuous}) \\
&\leq C\|z\|^2 - J'(\sigma_{m-1} + z) \cdot z_m \quad (\text{from } 2.23) \\
&= C\|z\|^2 - (J'(\sigma_{m-1} + z) - J'(\sigma_{m-1} + z_m)) \cdot z_m \\
&\leq C\left(\|z\|^2 + \|z - z_m\|\,\|z_m\|\right) \\
&\leq C\left(\|z\|^2 + \|z\|\,\|z_m\| + \|z_m\|^2\right).
\end{aligned}
$$

Hence, replacing z by $\pm z \dfrac{\|z_m\|}{\|z\|}$ in the previous inequalities, we obtain

$$|J'(\sigma_{m-1}) \cdot z| \leq 3C\|z_m\|\,\|z\|. \qquad\qquad \square$$

LEMMA 2.6. *With the same notations as Theorem 2.2, one has*

$$\lim\limits_{m \to +\infty} J'(\sigma_m) \cdot \sigma_m = 0,$$

and the sequence PGD $(\sigma_m)_m$ given by the scheme (M) converges to the solution of (2.8).

Proof. Applying the fact that $\sum\limits_{k=1}^{+\infty} \widetilde{J}'(z_k) \cdot z_k < +\infty$ and $\lim\limits_{m \to +\infty} J'(0) \cdot \sigma_m = J'(0) \cdot \widehat{\sigma}$, where $\widehat{\sigma}$ is the weak limit of $(\sigma_m)_{m \in \mathbb{N}}$ (and at the same time the unique solution of (2.8) from Theorem 2.2), we deduce, using (2.20), that the real sequence $(J'(\sigma_m) \cdot \sigma_m)_m$ is convergent.

On the other hand, one has

$$
\begin{aligned}
|J'(\sigma_m) \cdot \sigma_m| &\leq \sum_{k=1}^{m} |J'(\sigma_m) \cdot z_k| \\
&\leq C \sum_{k=1}^{m} \|z_{m+1}\| \, \|z_k\| \qquad \text{(Lemma 2.5)} \\
&\leq C \left(m \, \|z_{m+1}\|^2\right)^{1/2} \left(\sum_{k=1}^{m} \|z_k\|^2\right)^{1/2} \qquad \text{(Hölder's inequality)} \\
&\leq C \left((m+1) \, \|z_{m+1}\|^2\right)^{1/2} \left(\sum_{k=1}^{+\infty} \|z_k\|^2\right)^{1/2}
\end{aligned}
$$

Finally, as $\sum_{k=1}^{+\infty} \|z_k\|^2 < +\infty$ there is a subsequence $(z_{m_k})_k$ of $(z_m)_m$ such that $\lim_{k \to +\infty} m_k \|z_{m_k}\|^2 = 0$. Therefore,

$$
\lim_{m \to +\infty} J'(\sigma_m) \cdot \sigma_m = \lim_{k \to +\infty} J'(\sigma_{m_k}) \cdot \sigma_{m_k} = 0.
$$

Applying the fact that

$$
\begin{cases}
\sigma_m & \rightharpoonup \quad \sigma \quad \text{in } W_{\|\cdot\|} \\
J'(\sigma_m) \cdot \sigma_m & \to \quad 0 \quad \text{for } m \to +\infty
\end{cases}
$$

and the decomposition formulae (2.15) and (2.15), we deduce that

$$
J'(\sigma_m - \sigma) \cdot (\sigma_m - \sigma) \to 0 \quad \text{for } m \to +\infty.
$$

Consequently, we obtain

$$
(2.24) \qquad J'(0) \cdot (\sigma_m - \sigma) + \widetilde{J}'(\sigma_m - \sigma) \cdot (\sigma_m - \sigma) \to 0 \quad \text{for } m \to +\infty,
$$

which concludes the proof. $\qquad\qquad\square$

Remark *The novelty of our convergence proof is to bring out that weak convergence of the PGD sequence $(\sigma_m)_m$ is due to the Brezis-Lieb decomposition [7]. Our result can then lead to PGD weak convergence for a large class of variational problems where strong convergence can not hold true. On the other hand, in the literature ([14] and the references therein), the ellipticity condition (H_2) refers to J but in the present work it refers to the modified energy $\widetilde{J}$, that is, the weak-convergence overcomes the affine part of J.*

Another approach can be mentioned, the $L^2 - L^1$ one, with

$$
(2.25) \qquad J_0(\sigma) := \int_{\Omega} \left| \sigma(x,y) \cdot \nabla I(x,y) + \frac{\partial I}{\partial t}(x,y) \right| \mathrm{d}x \, \mathrm{d}y.
$$

and

(2.26) $$J(\sigma) := J_0(\sigma) + J_1(\sigma),$$

where

$$J_1(\sigma) := \int_\Omega c(x,y)|\sigma(x,y)|\, dx\, dy + \lambda \int_\Omega |\nabla\sigma(x,y)|^2\, dx\, dy.$$

If Ω is a bounded domain in $\mathbb{R}^2$ then using the continuous Sobolev imbeddings $W^{1,2}(\Omega) \subset W^{1,1}(\Omega) \subset L^2(\Omega)$, we remark that

$$\|u\|_{L^2(\Omega)} \le c_1 \left(\|u\|_{L^1(\Omega)} + \|\nabla u\|_{L^1(\Omega)} \right) \le c_2 \left(\|u\|_{L^1(\Omega)} + \|\nabla u\|_{L^2(\Omega)} \right), \quad \forall\, u \in W^{1,2}(\Omega)$$

and this second approach can be handled in the same way as the $L^2 - L^2$, provided make J differentiable. To perform this, we can replace $|s|$ by $\sqrt{s^2 + \varepsilon^2}$, with $s = \sigma(x,y) \cdot \nabla I(x,y) + \frac{\partial I}{\partial t}(x,y)$ or $s = \sigma(x,y)$ in the expression of J and ε sufficiently small.

However, the approach $L^1 - L^2$ can not be performed, since this requires the regularisation by total variation, and thus, in turn, the functional framework of the space of functions bounded variation, which is not a Banach Tensor space [25].

In what follows, we will implement the PGD method described in (M) in the two dimensional case. Then $\Omega = \Omega_1 \times \Omega_2$, and generic points of Ω will be denoted by (x,y). The planar OF will be denoted by $\sigma(x,y) = (u(x,y), v(x,y))$.

2.2. Implementation of tested OF approaches. Using quadratic penalizers, outliers (such as noise or occultations) may get high influence on the energy. It has been proposed in the past to employ non-convex penalization of the OF constraint of Eq. (2.6), in order to obtain a robust data term [6]. Unfortunately, such penalizations lead to ill-posed mathematical minimization problems. Two typical OF methods are described and evaluated. They differ in the penalizer applied on the data fidelity term $s = I_t + \nabla I \cdot \sigma$. A first one is based on a quadratic penalizer (referred to as "$L^2 - L^2$ approach", $R(s) = s^2$), and a second one on a L^1 regularization (referred to as "$L^2 - L^1$ approach", $R(s) = \sqrt{s^2 + \epsilon^2}$, with $\epsilon = 10^{-3}$). An experimental setup is then presented to compare implemented PGD methods to their corresponding non-PGD implementations in terms of accuracy of motion estimates as well as computation time.

2.2.1. General data processing sequence of the implemented OF approaches. The general implementation of the implemented OF approaches is outlined for both approaches as a sequence diagram in Fig. 1. We denote from now I_{ref} and I_{cur} as the reference and the current images to register. A multi-resolution scheme was performed which iterated the registration from a 64×64 downsampled image step by step to the original image resolution [24]. Common processing steps are reported in white boxes. Specific implementations of the processing step in grey are detailed in sections 2.2.2 and 2.2.3 for the PGD-based and the non-PGD approaches, respectively.
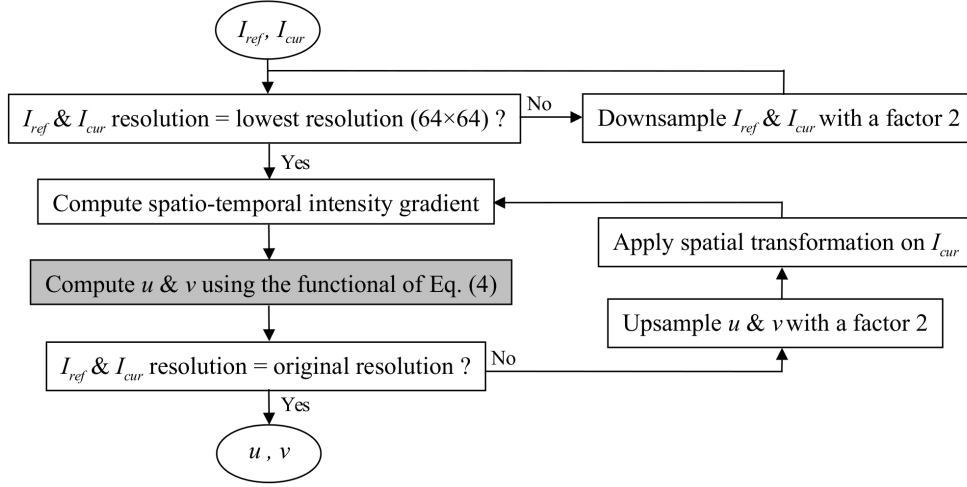
FIG. 1. *Data processing sequence of the implemented OF approaches. Specific implementations of the processing step in grey is detailed in sections 2.2.2 and 2.2.3 for the PGD-based and the non-PGD approaches, respectively.*

2.2.2. Proposed PGD-based approach. Applying the Euler-Lagrange equation on Eq. (2.8), one can derive two equations for each $(x, y) \in \Omega$, as follows

$$(2.27) \qquad \begin{cases} R'\left(I_x u + I_y v + I_t\right) I_x - \lambda \Delta u &= 0 \\ R'\left(I_x u + I_y v + I_t\right) I_y - \lambda \Delta v &= 0 \end{cases}$$

For the L^2 data term, one just gets $R'(s) = 2s$, whereas $R'(s) = s/R(s)$ for the L^1 data term. The system of Eq. (2.27) can therefore be rewritten as follows

$$(2.28) \qquad \begin{cases} I_x^2 u + I_x I_y v + I_x I_t - \lambda K(u,v)\Delta u &= 0 \\ I_x I_y u + I_y^2 v + I_y I_t - \lambda K(u,v)\Delta v &= 0 \end{cases}$$

with $K(u, v) = 1/2$ for the L^2 data term and

$$(2.29) \qquad K(u, v) = \sqrt{(I_x u + I_y v + I_t)^2 + \epsilon^2}.$$

for the L^1 one. The proposed solution strategy, based on three nested loops, is summarized in Fig. 2. Our implementation takes maximum benefit of terms that can be expressed linearly, leaving an explicit non-linear expression for the other terms.

The main loop (which iterates in m) seeks the solution of the following iterative process

$$(2.30) \qquad \begin{cases} u_m &= u_{m-1} + \varphi_m \otimes \psi_m \\ v_m &= v_{m-1} + \widetilde{\varphi}_m \otimes \widetilde{\psi}_m \end{cases}$$

with

$$(2.31) \qquad \begin{cases} u_{m-1} &= \sum_{i=1}^{m-1} \varphi_i \otimes \psi_i \\ v_{m-1} &= \sum_{i=1}^{m-1} \widetilde{\varphi}_i \otimes \widetilde{\psi}_i \end{cases}$$

$$I_{ref},\ I_{cur}$$

Set m to 1

Set l to 1

Set k to 1

Compute $\left(\varphi_m\right)_l^k$ and $\left(\widetilde{\varphi_m}\right)_l^k$ using Eq. (30)

Iterate k

$\left\| \left(\varphi_m, \widetilde{\varphi_m}\right)_l^k - \left(\varphi_m, \widetilde{\varphi_m}\right)_l^{k-1} \right\|_2 < \varepsilon\ ?$ No

Yes

Set k to 1

Iterate l Iterate m

Compute $\left(\psi_m\right)_l^k$ and $\left(\widetilde{\psi_m}\right)_l^k$ using Eq. (31)

Iterate k

$\left\| \left(\psi_m, \widetilde{\psi_m}\right)_l^k - \left(\psi_m, \widetilde{\psi_m}\right)_l^{k-1} \right\|_2 < \varepsilon\ ?$ No

Yes

$\left\| \left(\varphi_m \otimes \psi_m\right)_l^k - \left(\widetilde{\varphi_m} \otimes \widetilde{\psi_m}\right)_{l-1}^k \right\|_2 < \varepsilon\ ?$ No

Yes

$\left\| \left(u_m, v_m\right)_l^k \right\|_2 < \varepsilon\ ?$ No
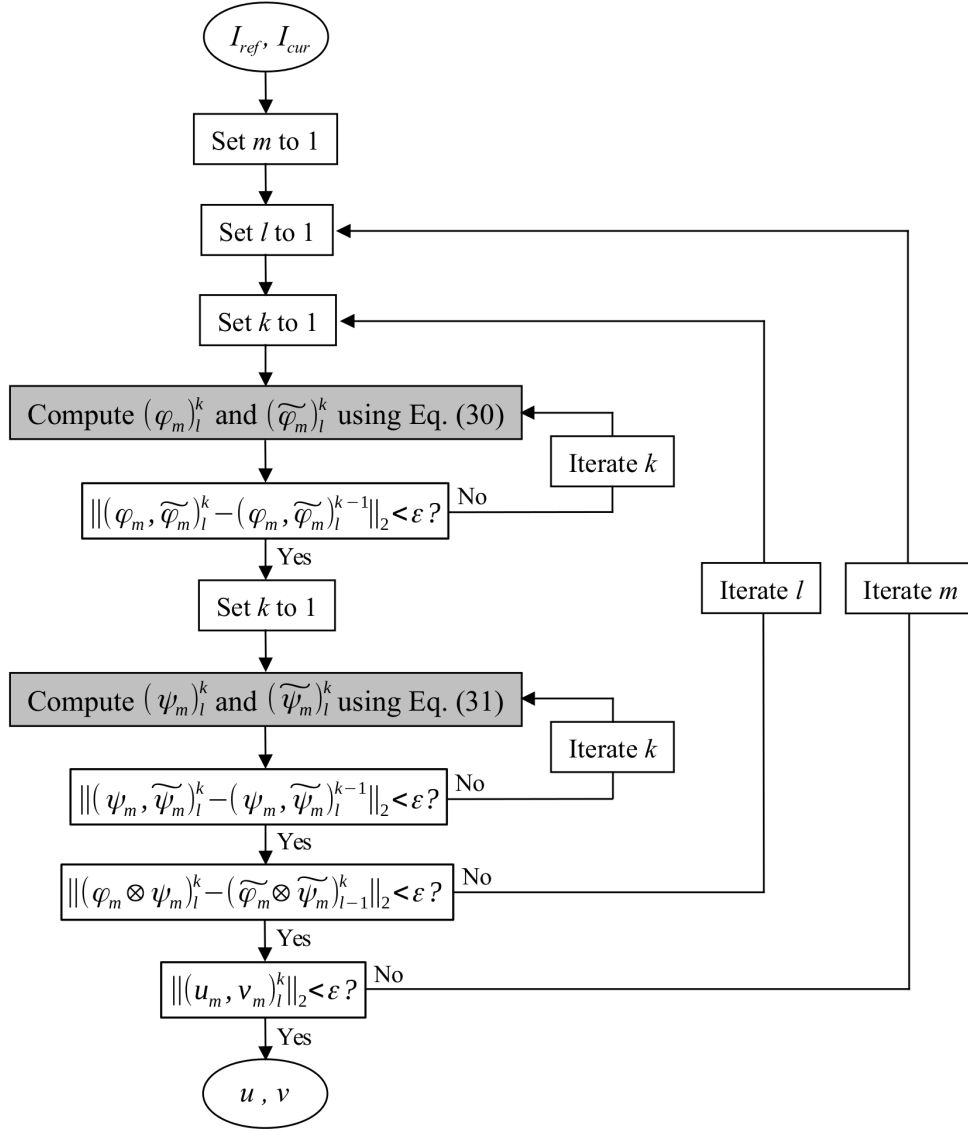
Yes

$$u\ ,\ v$$

FIG. 2. *Data processing sequence of the proposed PGD-based OF solution. The two steps in grey correspond to the most computationally intensive numerical tasks, the rest of the algorithm being mainly composed by conditional tests.*

Spatio-temporal derivative $I_{x,y\ ,t}$ were updated at each iteration (a centered finite difference scheme was employed for the computation of $I_{x,y}$) and the norm of the increment $\left\|(u,v)_m\right\|_2$ was compared to a maximal allowed tolerance ε in order to ensure the convergence. A specific ε value was employed for each level of the multi-resolution scheme described in section 2.2.1. While a starting value of $\varepsilon = 0.01$ was used for the original image resolution, this value was divided by a factor 2 each time I_{ref} and I_{cur} were down-sampled with a factor 2. That way, higher numbers of iterations were employed at low image resolution levels, the latter being less time

consuming.

For each iteration m, PGD components $\left(\varphi, \widetilde{\varphi}, \psi, \widetilde{\psi}\right)_m$ are the solution of the following non-linear system of $N_1^2 \times N_2^2$ equations (we recall that N_1 and N_2 denote the number of pixels along the horizontal and the vertical directions, respectively).

$$
(2.32) \quad
\begin{cases}
\begin{aligned}
I_x^2 \varphi_m \psi_m \; &+ \; I_x I_y \widetilde{\varphi}_m \widetilde{\psi}_m \\
&- \; \lambda K(u_m, v_m) \varphi_m'' \psi_m \\
&- \; \lambda K(u_m, v_m) \varphi_m \psi_m'' \\
= \; &- \; I_x I_t \\
&+ \; \lambda K(u_m, v_m) \Delta u_{m-1} \\
I_x I_y \varphi_m \psi_m \; &+ \; I_y^2 \widetilde{\varphi}_m \widetilde{\psi}_m \\
&- \; \lambda K(u_m, v_m) \widetilde{\varphi}_m'' \widetilde{\psi}_m \\
&- \; \lambda K(u_m, v_m) \widetilde{\varphi}_m \widetilde{\psi}_m'' \\
= \; &- \; I_y I_t \\
&+ \; \lambda K(u_m, v_m) \Delta v_{m-1}
\end{aligned}
\end{cases}
$$

That way, it must be underlined that the data fidelity term relies on updated variables at each iteration m, which reduces the impact of the Taylor approximation introduced in Eq. (2.6). This means that I_t, I_x and I_y are registered with respect to u_{m-1} and v_{m-1}. The regularization, for its part, is performed of the overall resulting motion.

We divide this task in two successive problems, one for the calculation of $(\varphi, \widetilde{\varphi})_m$ and one for the calculation of $\left(\psi, \widetilde{\psi}\right)_m$. The solving of these two problems is repeated iteratively (see the second nested loop, which iterates in l, in Fig. 2) until the residual $\left\| \left(\widetilde{\varphi}_m \otimes \widetilde{\psi}_m\right)_l - \left(\widetilde{\varphi}_m \otimes \widetilde{\psi}_m\right)_{l-1} \right\|_2$ reaches a user-defined tolerance ε. The ε value employed for the l-loop was similar to the one employed for the m-loop detailed above.

First components of the tensorial product $(\varphi_m, \widetilde{\varphi}_m)_l$ were computed as follows: Multiplying the first and the second equation of Eq. (2.32) by $(\psi_m)_l$ and $\left(\widetilde{\psi}_m\right)_l$, respectively, and integrating the result along the y direction, one can derive the following system of $2 \times N_1$ equations ($(\varphi_m)_l(x)$ and $(\widetilde{\varphi}_m)_l(x)$ being the unknowns)

$$
(2.33) \quad
\begin{cases}
A(x)\,(\varphi_m'')_l(x) + B(x)\,(\varphi_m)_l(x) + C(x)\,(\widetilde{\varphi}_m)_l(x) &= D(x) \\
\widetilde{A}(x)\,(\widetilde{\varphi}_m'')_l(x) + \widetilde{B}(x)\,(\varphi_m)_l(x) + \widetilde{C}(x)\,(\widetilde{\varphi}_m)_l(x) &= \widetilde{D}(x)
\end{cases}
$$

where

$$
\begin{cases}
A(x) &=& -\lambda \int_{\Omega_2}[K\left(u_m(x,y),v_m(x,y)\right)(\psi_m)_l\,(y)^2]\mathrm{d}y \\
B(x) &=& \lambda \int_{\Omega_2}[\partial_y K\left(u_m(x,y),v_m(x,y)\right)(\psi_m)_l\,(y)\,(\psi'_m)_l\,(y)]\,\mathrm{d}y \\
&& +\lambda \int_{\Omega_2}\left[K\left(u_m(x,y),v_m(x,y)\right)(\psi'_m)_l\,(y)^2\right]\mathrm{d}y \\
&& +\int_{\Omega_2}\left[I_x(x,y)^2\,(\psi_m)_l\,(y)^2\right]\mathrm{d}y \\
C(x) &=& \int_{\Omega_2}[I_x(x,y)I_y(x,y)\left(\widetilde{\psi}_m\right)_l(y)\,(\psi_m)_l\,(y)]\mathrm{d}y \\
D(x) &=& \lambda \int_{\Omega_2}[K\left(u_m(x,y),v_m(x,y)\right)\Delta u_{m-1}(x,y)\,(\psi_m)_l\,(y)]\mathrm{d}y \\
&& -\int_{\Omega_2}[I_x(x,y)I_t(x,y)\,(\psi_m)_l\,(y)]\mathrm{d}y \\
\widetilde{A}(x) &=& -\lambda \int_{\Omega_2}[K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\psi}_m\right)_l(y)^2]\mathrm{d}y \\
\widetilde{B}(x) &=& C(x) \\
\widetilde{C}(x) &=& \lambda \int_{\Omega_2}\left[\partial_y K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\psi}_m\right)_l(y)\left(\widetilde{\psi}'_m\right)_l(y)\right]\mathrm{d}y \\
&& +\lambda \int_{\Omega_2}\left[K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\psi}'_m\right)_l(y)^2\right]\mathrm{d}y \\
&& +\int_{\Omega_2}\left[I_y(x,y)^2\left(\widetilde{\psi}_m\right)_l(y)^2\right]\mathrm{d}y \\
\widetilde{D}(x) &=& \lambda \int_{\Omega_2}[K\left(u_m(x,y),v_m(x,y)\right)\Delta v_{m-1}(x,y)\left(\widetilde{\psi}_m\right)_l(y)]\mathrm{d}y \\
&& -\int_{\Omega_2}[I_y(x,y)I_t(x,y)\left(\widetilde{\psi}_m\right)_l(y)]\mathrm{d}y
\end{cases}
$$

Note that an integration by part was performed in $B(x)$ and $\widetilde{C}(x)$ in order to remove the second order derivative expressions of ψ_m and $\widetilde{\psi}_m$, respectively. From here, a linear iterative scheme, based on the Jacobi method, has been set up as follows (see the first of the two nested loops iterating along k, in Fig. 2)

- Second order terms Δu_{m-1} and Δv_{m-1} were computed explicitly as follows: An approximation of the Laplacian was used, as suggested by Horn&Schunck, where $\Delta u_{m-1} = (\overline{u}_{m-1} - u_{m-1})$, $\overline{u}_{m-1}$ being the mean value of u_{m-1} in a neighborhood (3×3 pixels) of the estimated point [17] (the same approximation was employed for the calculation of Δv_{m-1}).
- The non-linear term for the L^1 data expression $K\left(u_m,v_m\right)$ was computed explicitly as follows

$$(2.34) \qquad K\left(u_m,v_m\right) = \sqrt{(I_x u_{m-1} + I_y v_{m-1} + I_t)^2 + \epsilon^2}.$$

- Second derivative terms $\left(\varphi''_m, \widetilde{\varphi}''_m, \psi''_m, \widetilde{\psi}''_m\right)_l$ were linearized as follows: At each iteration k, we used the approximation $(\varphi''_m)_l^k = \left(\overline{(\varphi_m)_l}^{k-1} - (\varphi_m)_l^k\right)$, $\overline{(\varphi_m)_l}^{k-1}$ being the mean value of $(\varphi_m)_l^{k-1}$ in a neighborhood (3 pixels) of the estimated point. The same approximation was employed for the calculation of the other terms $\left(\widetilde{\varphi}''_m, \psi''_m, \widetilde{\psi}''_m\right)_l$. That way, $\left(\overline{\varphi_m}, \overline{\widetilde{\varphi}_m}, \overline{\psi_m}, \overline{\widetilde{\psi}_m}\right)_l$ were expressed explicitly, leaving an implicit linear formulation in $(\varphi_m)_l^k$ and $(\widetilde{\varphi}_m)_l^k$.

Solutions $(\varphi_m, \widetilde{\varphi}_m)_l$ of Eq. (2.33) could thus be obtained using the following fixed

point schemes

$$(2.35) \quad \begin{cases} (\varphi_m)_l^k & = & \dfrac{\left(D - A\overline{(\varphi_m)_l}^{k-1}\right)\left(\widetilde{C} - \widetilde{A}\right) - \left(\widetilde{D} - \widetilde{A}\overline{(\widetilde{\varphi}_m)_l}^{k-1}\right)C}{(B-A)\left(\widetilde{C}-\widetilde{A}\right)-\widetilde{B}C} \\[2ex] (\widetilde{\varphi}_m)_l^k & = & \dfrac{(B-A)\left(\widetilde{D}-\widetilde{A}\overline{(\widetilde{\varphi}_m)_l}^{k-1}\right) - \widetilde{B}\left(D-A\overline{(\varphi_m)_l}^{k-1}\right)}{(B-A)\left(\widetilde{C}-\widetilde{A}\right)-\widetilde{B}C} \end{cases}$$

The residual $\left\|(\varphi_m,\widetilde{\varphi}_m)_l^k - (\varphi_m,\widetilde{\varphi}_m)_l^{k-1}\right\|_2$ was compared to a maximal allowed tolerance $\varepsilon = 10^{-4}$ in order to ensure the convergence.

It is important to underline that its is only necessary to solve a system containing $2 \times N_1$ equations in order to get $(\varphi_m, \widetilde{\varphi}_m)$.

The second components of the tensorial product $\left(\psi_m, \widetilde{\psi}_m\right)_l^k$ were then obtained identically (see the second of the two nested loops iterating along k, Fig. 2), using however the updated $(\varphi_m, \widetilde{\varphi}_m)_l$, as follows

$$(2.36) \quad \begin{cases} (\psi_m)_l^k & = & \dfrac{\left(H - E\overline{(\psi_m)_l}^{k-1}\right)\left(\widetilde{G} - \widetilde{E}\right) - \left(\widetilde{H} - \widetilde{E}\overline{(\widetilde{\psi}_m)_l}^{k-1}\right)G}{(F-E)\left(\widetilde{G}-\widetilde{E}\right)-\widetilde{F}G} \\[2ex] \left(\widetilde{\psi}_m\right)_l^k & = & \dfrac{(F-E)\left(\widetilde{H}-\widetilde{E}\overline{(\widetilde{\psi}_m)_l}^{k-1}\right) - \widetilde{F}\left(H-E\overline{(\psi_m)_l}^{k-1}\right)}{(F-E)\left(\widetilde{G}-\widetilde{E}\right)-\widetilde{F}G} \end{cases}$$

where

$$\begin{cases} E(y) & = & -\lambda \int_{\Omega_1}[K\left(u_m(x,y),v_m(x,y)\right)\left(\varphi_m\right)_l(x)^2]\mathrm{d}x \\ F(y) & = & \lambda \int_{\Omega_1}\left[\partial_x K\left(u_m(x,y),v_m(x,y)\right)\left(\varphi_m\right)_l(x)\left(\varphi'_m\right)_l(x)\right]\mathrm{d}x \\ & & +\lambda \int_{\Omega_1}\left[K\left(u_m(x,y),v_m(x,y)\right)\left(\varphi'_m\right)_l(x)^2\right]\mathrm{d}x \\ & & +\int_{\Omega_1}\left[I_x(x,y)^2\left(\varphi_m\right)_l(x)^2\right]\mathrm{d}x \\ G(y) & = & \int_{\Omega_1}[I_x(x,y)I_y(x,y)\left(\widetilde{\varphi}_m\right)_l(x)\left(\varphi_m\right)_l(x)]\mathrm{d}x \\ H(y) & = & \lambda \int_{\Omega_1}[K\left(u_m(x,y),v_m(x,y)\right)\Delta u_{m-1}(x,y)\left(\varphi_m\right)_l(x)]\mathrm{d}x \\ & & -\int_{\Omega_1}\left[I_x(x,y)I_t(x,y)\left(\varphi_m\right)_l(x)\right]\mathrm{d}x \\ \widetilde{E}(y) & = & -\lambda \int_{\Omega_1}[K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\varphi}_m\right)_l(x)^2]\mathrm{d}x \\ \widetilde{F}(y) & = & G(y) \\ \widetilde{G}(y) & = & \lambda \int_{\Omega_1}\left[\partial_x K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\varphi}_m\right)_l(x)\left(\widetilde{\varphi}'_m\right)_l(x)\right]\mathrm{d}x \\ & & +\lambda \int_{\Omega_1}\left[K\left(u_m(x,y),v_m(x,y)\right)\left(\widetilde{\varphi}'_m\right)_l(x)^2\right]\mathrm{d}x \\ & & +\int_{\Omega_1}\left[I_y(x,y)^2\left(\widetilde{\varphi}_m\right)_l(x)^2\right]\mathrm{d}x \\ \widetilde{H}(y) & = & \lambda \int_{\Omega_1}[K\left(u_m(x,y),v_m(x,y)\right)\Delta v_{m-1}(x,y)\left(\widetilde{\varphi}_m\right)_l(x)]\mathrm{d}x \\ & & -\int_{\Omega_1}\left[I_y(x,y)I_t(x,y)\left(\widetilde{\varphi}_m\right)_l(x)\right]\mathrm{d}x \end{cases}$$

Again, one can notice that another set of only $2 \times N_2$ equations was necessary in order to estimate $\left(\psi_m, \widetilde{\psi}_m\right)_l^k$.

2.2.3. Implementation of the corresponding non-PGD approach. This section describes the existing non-PGD implementation employed to assess the benefit of the proposed PGD approach in terms of accuracy of motion estimates as well as computation time. To clarify the advantages on the final results, our non-PGD

implementation takes benefit of terms that can be expressed linearly, leaving an explicit non-linear expression for the others, similarly as previously. For both methods, the same convergence criteria were employed. Similar implementations were also employed for the computation of the spatio-temporal derivatives and the approximation of the Laplacian. The reader is referred to the section 2.2.2 for more details on the implementation.

The solving strategy of the non-PGD approach, based on two nested loops, is summarized in Fig. 3.
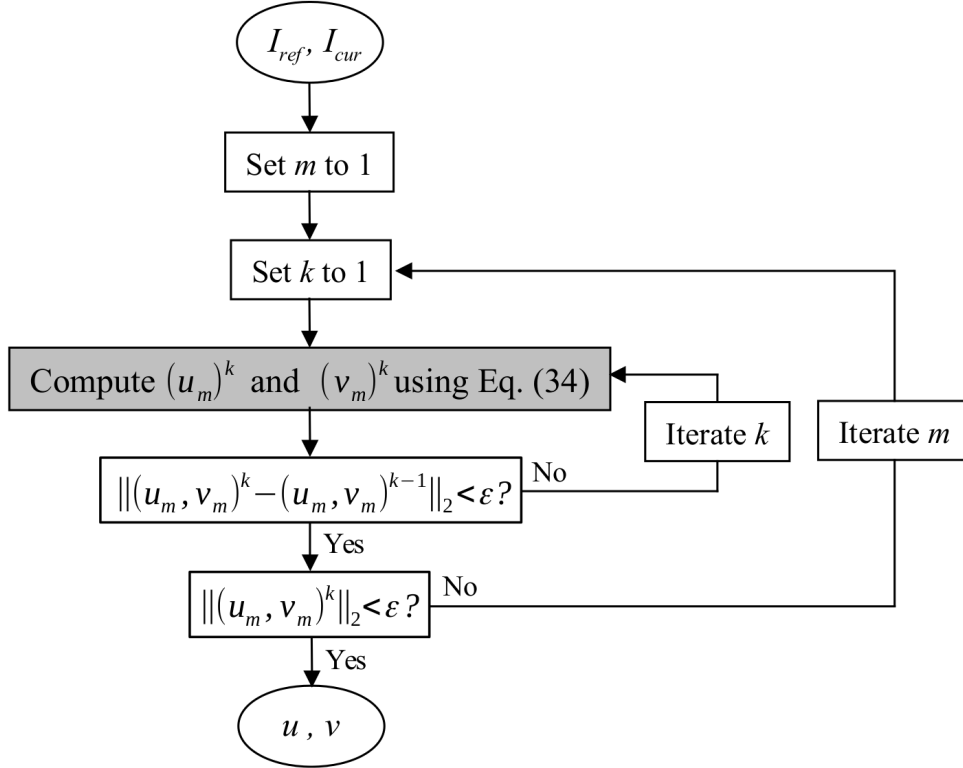

FIG. 3. *Data processing sequence of the non-PGD OF solution. The step in grey corresponds to the most computationally intensive numerical task, the rest of the algorithm being mainly composed by conditional tests.*

The main loop (which iterates in m) is an iterative refinement which updates $I_{x,y,t}$ within each resolution. The solution of the following iterative process was seek

$$(2.37) \qquad \begin{cases} u &= \sum_{m=1}^{\infty} u_m \\ v &= \sum_{m=1}^{\infty} v_m \end{cases}$$

To ensure the convergence of the algorithm, the norm of the increment $\|(u,v)_m\|_2$ was compared to a maximal allowed tolerance ε. The employed ε value was identical to the one used for the PGD case (see section 2.2.2). From the system of Eq. (2.28), one have that partial motion field components $(u,v)_m$ are solution of the following

non-linear system of $N_1^2 \times N_2$ equations

$$(2.38) \quad \begin{cases} \begin{array}{llll} I_x^2 & u_m & + & I_x I_y \quad\quad\quad v_m \\ & & - & \lambda K(u_m, v_m) \quad \Delta u_m \quad = \quad -I_x I_t + \lambda K(u_m, v_m) \quad \Delta u_{m-1} \\ I_x I_y & u_m & + & I_y^2 \quad\quad\quad\quad v_m \\ & & - & \lambda K(u_m, v_m) \quad \Delta v_m \quad = \quad -I_y I_t + \lambda K(u_m, v_m) \quad \Delta v_{m-1} \end{array} \end{cases}$$

That way, similar to the proposed PGD implementation, we recall that the data fidelity term relies on updated variables at each iteration m, which reduces the impact of the Taylor approximation introduced in Eq. (2.6). The regularization, for its part, is performed of the overall resulting motion (*i.e.* $(u_m + u_{m-1}, v_m + v_{m-1})$).

Whithin each iteration m, the Euler-Lagrange equations of Eq. (2.28) were solved using an iterative scheme based on the Jacobi method (see the second nested loop, which iterates in k, Fig. 3) as follows

$$(2.39)$$
$$\begin{cases} (u_m)^k = \overline{(u_m)}^{k-1} + \Delta u_{m-1} - I_x \left(\dfrac{I_x\left(\overline{(u_m)}^{k-1} + \Delta u_{m-1}\right) + I_y\left(\overline{(v_m)}^{k-1} + \Delta v_{m-1}\right) + I_t}{I_x^2 + I_y^2 + \lambda K(u_m, v_m)} \right) \\ (v_m)^k = \overline{(v_m)}^{k-1} + \Delta v_{m-1} - I_y \left(\dfrac{I_x\left(\overline{(u_m)}^{k-1} + \Delta u_{m-1}\right) + I_y\left(\overline{(v_m)}^{k-1} + \Delta v_{m-1}\right) + I_t}{I_x^2 + I_y^2 + \lambda K(u_m, v_m)} \right) \end{cases}$$

The residual $\left\| (u_m, v_m)^k - (u_m, v_m)^{k-1} \right\|_2$ was compared to a maximal allowed tolerance $\varepsilon = 10^{-4}$ in order to ensure the convergence.

At this point, it is interesting to note that it is required to solve a system of $N_1 \times N_2$ equations in order to get u_m and v_m.

2.2.4. Extension to color images. A higher level of information is available on color images since for each pixel, one piece of information stems from each color channel (*i.e.* red, green and blue). The functional of Eq. (2.8), which only accounts for a single channel, was generalized for color images using J_0 rewritten as follows

$$(2.40) \quad\quad\quad J_0 = \int_\Omega \left[\sum_{c=1}^{3} R\left(\nabla I_c(x, y) + \frac{\partial I}{\partial t}(x, y) \right) \right] \mathrm{d}x\mathrm{d}y,$$

where c is an index iterating over the red/grey/blue color channels. The reader is referred to section 2.2 for the determination of R.

The generalization of numerical schemes associated to the PGD and the non-PGD approaches (described in sections 2.2.2 and 2.2.3, respectively) is then rather straightforward.

2.3. Experimental setup. The ability of the proposed PGD-based approach to solve $L^2 - L^2$ and $L^2 - L^1$ OF problems on images containing various amount of information (in terms of image resolution and/or SNR) was assessed on the Middlebury data sets [4]. The tested Middlebury collection includes eight data sets composed by a reference image, images to register and the associated groundtruth motion field. Both greyscale and color versions for the reference and the image to register are available.

The performance of the PGD-based approach was compared to the non-PGD implementation for images undergoing the following experimental conditions:

- To investigate the impact of the image resolution on the performance of the proposed PGD method, various data dimensions (*i.e.* the reference image, the image to register and the associated goldstandard displacement field) were constructed using a bicubic interpolation. Image resolutions of 64×64 and 128×128 were thus generated.
- To investigate the impact of low SNR, both the reference and the image to register were disturbed by a simulated Gaussian noise prior to the motion estimation process. For this purpose, a noise standard deviation equal 5% of the maximum color channel intensity were employed. Registration was then performed on a pair of images affected by the same level, but resulting from different realizations of the noise.
- OF algorithms are known to perform better on pre-filtered images. The reason is the following: The OF implementation estimates the displacement of a pixel only locally without taking into account any interaction between neighboring pixels. Consequently, it may run into problems once the gradient disappears somewhere, or if only the flow in normal direction to the gradient can be estimated (known in the literature as the "aperture problem" [1]). In addition, some outliers may be expected in the estimates. It is consequently useful to introduce as a further assumption the smoothness of the flow field. To investigate the impact of the image pre-filtering, both the reference and the image to register were filtered by a spatial Gaussian filter prior the motion estimation process, as described in [5]. For this purpose, a 5×5 Gaussian kernel with a standard deviation equal to 0.3 was employed.
- Since a higher level of information is available on color images (for each pixel, one piece of information stems from each color channel), the proposed PGD-based method was evaluated on both greyscale and color images.

During the registration process, the intensity of each channel image were normalized such that they lie between 0 and 1. Since in this interval $|x| \geq x^2$, it is expected that the λ value with the $L^2 - L^1$ functionals will need to be higher than the one employed by the $L^2 - L^2$ functionals in order to provide similar motion estimates. Arbitrarily chosen λ values of 0.1 and 0.5 were consequently employed in the scope of this study for the $L^2 - L^2$ and the $L^2 - L^1$ approaches, respectively.

Our test platform was an Intel Core CPU i3-M370 dual-core 2.40 GHz with 2 GB of RAM.

2.4. Assessement of the proposed PGD-based approach. The accuracy of each motion estimate was quantified on pixel-by-pixel basis using the Endpoint Error (EE) and the Angular Error (AE). Averaged EE and AE were compared for the existing non-PGD and the proposed PGD-based approaches for each tested data set and each experimental condition described above. For this purpose, the improvement ratio of the averaged EE and AE between the non-PGD and the proposed PGD-based approaches was computed (an improvement ratio higher than 1 tells that the proposed PGD-based approach outperforms the non-PGD one). The percentage of improved Middlebury data sets as well as the required computation times were also reported.

3. Results. Fig. 4 shows typical findings of the efficiency of the PGD-based OF obtained on a subset of three greyscale Middlebury video sequences, namely "Dimetrodon", "Grove 2" and "Hydrangea". For this purpose, the $L^1 - L^2$ functional was optimized on each couple of images, each one being resized to a resolution of 64×64 pixels. For each sequence, the error of motion estimates were found to be reduced using the PGD approach as compared to the non-PGD case (see the corre-
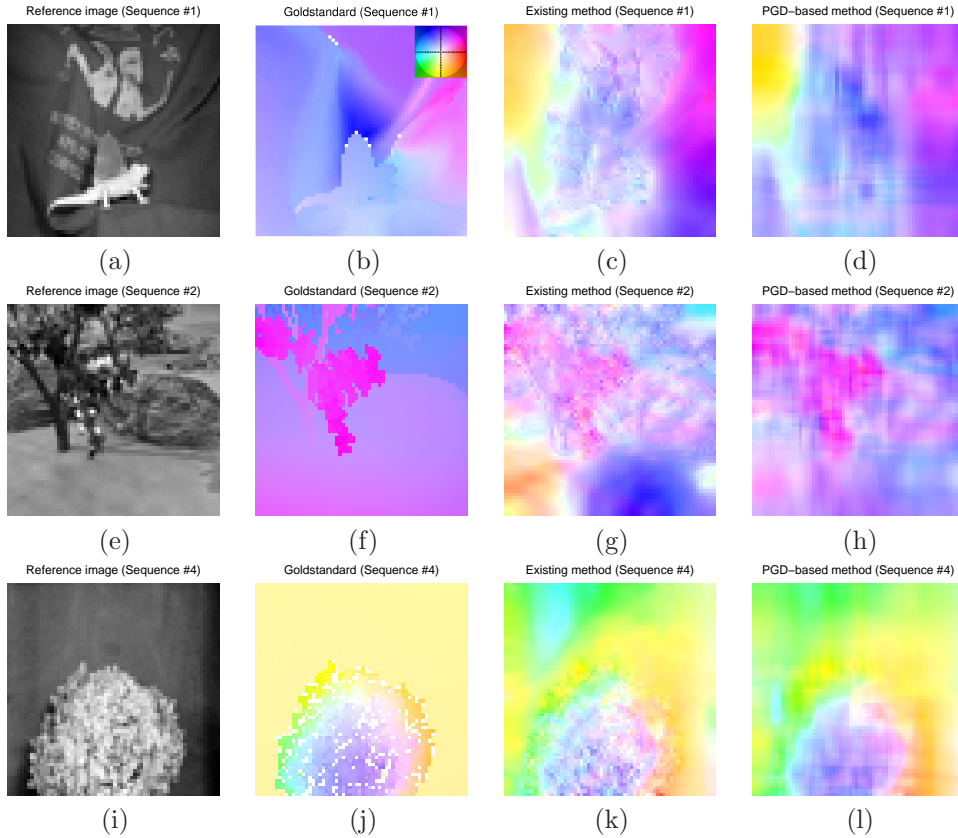
Fɪɢ. 4. *Typical findings of the efficiency of the PGD-based OF obtained on a subset of greyscale Middlebury video sequences using an image resolution of* 64×64 *pixels. These results were generated using the* $L^1 - L^2$ *functional. First column: tested images, second column: color coded gold-standard motion fields (the color scheme used to represent the motion fields is reported in the upper-right corner of (b)), third column: color coded motion fields estimated using the non-PGD OF approach, fourth column: color coded motion fields estimated using the proposed PGD-based OF approach.*

sponding pixelwise EE and AE error maps in Fig. 5). Table 1 summarizes results obtained for all tested Middlebury video sequences using the $L^1 - L^2$ functional and images of resolution 64×64 pixels. In average over the eight tested sequences, both endpoint and angular errors were improved by $\approx$25%, together with a reduction of the required computation time by a factor close to 6. When the SNR was decreased, performances in terms of both EE, AE and computation time were improved in all tested sequences using the PGD-based approach, as shown in Table 2. This was achieved in combination with a further improvement in terms of computation time (an averaged improvement ratio of 15 was obtained).

Each Table 3, 4 and 5 provides a complete performance summary obtained using greyscale images, pre-filtered greyscale images and color images, respectively. Each table reports results obtained under various resolution/functional/SNR conditions using both the non-PGD and the proposed PGD-based OF approach: Table 3 shows that, using greyscale 64×64 images, the proposed PGD-based approach outperformed the non-PGD approach in all tested conditions. It can also be observed that perfor-
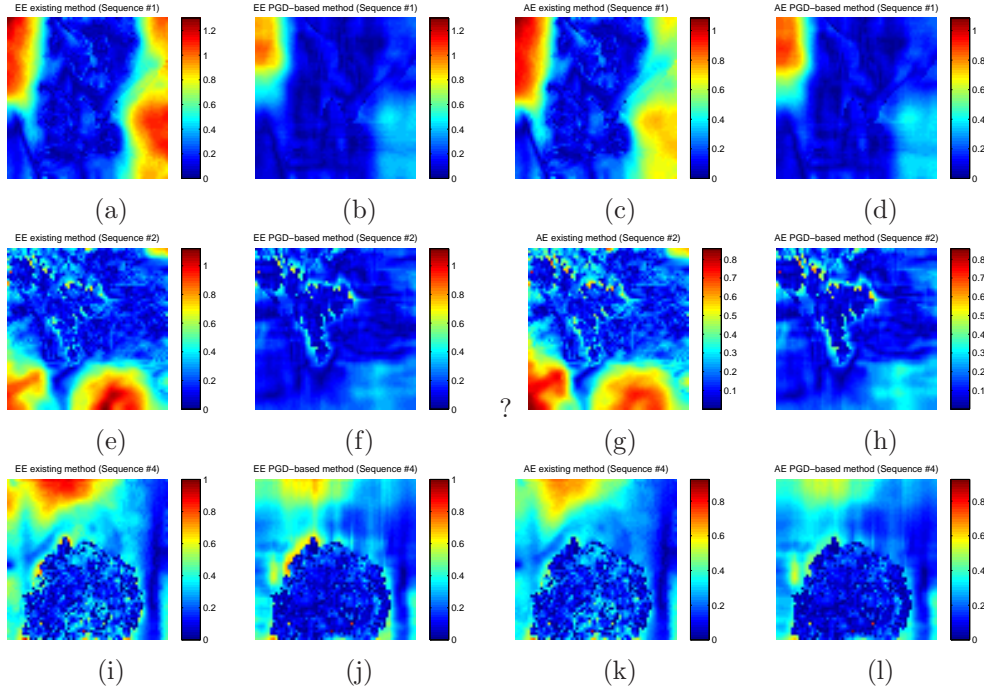
FIG. 5. *Pixelwise errors of the motion estimates reported in Fig. 4. First column: pixelwise endpoint errors obtained using the non-PGD approach, second column: pixelwise endpoint errors obtained using the proposed PGD-based approach, third column: pixelwise angular errors obtained using the non-PGD approach, fourth column: pixelwise angular errors obtained using the proposed PGD-based approach.*

Sequence Id	Middlebury sequence name	EE (pixel)		AE (radian)		Time (s)	
		non-PGD	PGD	non-PGD	PGD	non-PGD	PGD
#1	Dimetrodon	0.40	0.21	0.33	0.20	0.44	0.05
#2	Grove 2	0.32	0.16	0.28	0.15	0.51	0.09
#3	Grove 3	0.25	0.21	0.21	0.17	0.39	0.07
#4	Hydrangea	0.32	0.26	0.26	0.22	0.51	0.13
#5	RubberWhale	0.15	0.11	0.15	0.11	0.10	0.03
#6	Urban 2	0.17	0.17	0.12	0.12	0.29	0.07
#7	Urban 3	0.36	0.31	0.25	0.21	0.58	0.09
#8	Venus	0.18	0.23	0.14	0.18	0.48	0.06
Averaged Improvement Ratio (AIR)		1.31		1.27		5.73	
Pourcentage of Improved Sequences (PIS)		87.5%		87.5%		100%	

TABLE 1

Performances obtained using the non-PGD and the PGD-based OF approaches on the tested greyscale Middlebury sequences for an image resolution of 64×64 pixels using the $L^1 - L^2$ functional. In average over the eight tested Middlebury sequences, both endpoint and angular errors were improved by $\approx 25\%$, together with a reduction of the required computation time by a factor close to 6.

Sequence	Middlebury	EE (pixel)		AE (radian)		Time (s)	
Id	sequence name	non-PGD	PGD	non-PGD	PGD	non-PGD	PGD
#1	Dimetrodon	0.29	0.26	0.27	0.23	0.21	0.09
#2	Grove 2	0.28	0.26	0.25	0.22	0.33	0.10
#3	Grove 3	0.31	0.27	0.27	0.22	7.20	0.09
#4	Hydrangea	0.49	0.38	0.34	0.28	1.64	0.15
#5	RubberWhale	0.27	0.19	0.26	0.18	0.40	0.05
#6	Urban 2	0.31	0.29	0.21	0.19	0.72	0.09
#7	Urban 3	0.42	0.40	0.28	0.26	0.52	0.12
#8	Venus	0.30	0.28	0.25	0.23	0.47	0.06
Averaged Improvement Ratio (AIR)		1.17		1.18		15.00	
Pourcentage of Improved Sequences (PIS)		100%		100%		100%	

TABLE 2

Performances obtained using the non-PGD and the PGD-based OF approaches on the tested greyscale Middlebury sequences for an image resolution of 64×64 pixels using the $L^1 - L^2$ functional. Images were previously disturbed by a simulated Gaussian noise (noise standard deviation equal to 0.3, noise amplitude equal to 5% of the maximum grey level intensity). Performances in terms of both EE, AE and computation time were in this case improved in all tested sequences using the PGD-based approach.

Image	OF	Added	EE		AE		Time	
resolution	functional	noise	AIR	PIS	AIR	PIS	AIR	PIS
64×64	$L^2 - L^2$	No	1.00	37.5 %	1.00	37.5 %	6.93	100.0%
		Yes	1.32	100.0 %	1.29	100.0 %	4.56	100.0%
	$L^1 - L^2$	No	1.31	87.5 %	1.27	87.5 %	5.73	100.0%
		Yes	1.17	100.0 %	1.18	100.0 %	15.00	100.0%
128×128	$L^2 - L^2$	No	*0.89*	*12.5 %*	*0.90*	*12.5 %*	*6.96*	*100.0%*
		Yes	1.23	87.5 %	1.23	87.5 %	5.27	100.0%
	$L^1 - L^2$	No	*0.97*	*37.5 %*	*0.97*	*37.5 %*	*7.62*	*100.0%*
		Yes	1.09	87.5 %	1.08	87.5 %	7.49	100.0%

TABLE 3

Summary of the performances obtained for various resolution/functional/SNR conditions using the proposed PGD-based OF approach, as compared to the non-PGD OF approach. Results were here generated using greyscale images. The results are reported for image resolutions of 64×64 and 128×128 using both $L^2 - L^2$ and $L^1 - L^2$ functionals. The table also reports performances obtained using images previously disturbed by a simulated Gaussian noise (noise standard deviation equal to 0.3, noise amplitude equal to 5% of the maximum grey level intensity). Lines of the table corresponding to data-sets for which the PGD-based approach provided worse EE performances than the non-PGD one are reported using italic characters.

mances in terms of EE and AE were improved especially when using the $L^1 - L^2$ norm and/or noisy data sets. However, using noise-free images of resolution 128×128 pixels, the non-PGD approach outclassed the PGD approach in terms of motion estimates accuracy. In noisy data sets, the PGD-based approach outperformed the non-PGD approach for image resolutions of both 64×64 and 128×128 pixels. Table 4 shows that, using pre-filtered greyscale images, similar behaviors were observed for both the non-PGD and the PGD approaches as compared to the scenario without image

Image	OF	Added	EE		AE		Time	
resolution	functional	noise	AIR	PIS	AIR	PIS	AIR	PIS
64×64	$L^2 - L^2$	No	1.00	37.5 %	1.00	37.5 %	5.04	100.0%
		Yes	1.16	87.5 %	1.14	87.5 %	3.07	100.0%
	$L^1 - L^2$	No	1.12	37.5 %	1.10	37.5 %	4.97	100.0%
		Yes	1.25	100.0 %	1.25	100.0 %	4.63	100.0%
128×128	$L^2 - L^2$	No	*0.94*	*25.0 %*	*0.93*	*25.0 %*	*3.61*	*100.0%*
		Yes	1.03	50.0 %	1.04	50.0 %	3.87	100.0%
	$L^1 - L^2$	No	*0.91*	*25.0 %*	*0.93*	*25.0 %*	*5.07*	*100.0%*
		Yes	1.11	87.5 %	1.09	87.5 %	6.58	100.0%

TABLE 4

Summary of the performances obtained for various resolution/functional/SNR conditions using the proposed PGD-based OF approach, as compared to the non-PGD OF approach. Again, results were generated using greyscale images. However, both the image to register and the reference one were here spatially pre-filtered before the motion estimation process (for this purpose, a 5×5 Gaussian kernel with a standard deviation equal to 0.3 was employed). Lines of the table corresponding to data-sets for which the PGD-based approach provided worse EE performances than the non-PGD one are reported using italic characters.

Image	OF	Added	EE		AE		Time	
resolution	functional	noise	AIR	PIS	AIR	PIS	AIR	PIS
64×64	$L^2 - L^2$	No	1.04	37.5 %	1.04	37.5 %	3.29	100.0%
		Yes	1.16	87.5 %	1.15	87.5 %	2.85	100.0%
	$L^1 - L^2$	No	1.21	75.0 %	1.18	75.0 %	4.24	100.0%
		Yes	1.27	100.0 %	1.25	100.0 %	3.67	100.0%
128×128	$L^2 - L^2$	No	*0.91*	*12.5 %*	*0.90*	*12.5 %*	*7.19*	*100.0%*
		Yes	1.00	37.5 %	0.99	37.5 %	5.18	100.0%
	$L^1 - L^2$	No	*0.90*	*25.0 %*	*0.92*	*25.0 %*	*4.24*	*100.0%*
		Yes	1.02	62.5 %	1.04	62.5 %	6.69	100.0%

TABLE 5

Summary of the performances obtained for various resolution/functional/SNR conditions using the proposed PGD-based OF approach, as compared to the non-PGD OF approach. Here, the tested OF algorithms were run on color images. Lines of the table corresponding to data-sets for which the PGD-based approach provided worse EE performances than the non-PGD one are reported using italic characters.

pre-filtering reported in Table 3.

4. Discussion. As we mentioned earlier, PGD approaches are usually employed by the mechanical community to accelerate and even assess the solution of multidimensional models [2]. In the context of OF problems, our results interestingly show that the proposed PGD-based approach is rather beneficial for low image resolution and/or low SNR:

First, the gain of accuracy is explained as follows: Using the proposed PGD-based approach, each complete displacement field is fully described by two vectors, each one having a resolution equal to twice the horizontal and vertical image dimensions, respectively. In the proposed implementation, this data reduction was achieved using data integrations performed along horizontal and vertical spatial directions which are reported in Eq. (2.33) and (2.36). That way, the displacement of a pixel is in a sense more "global" since it accounts for the data fidelity and the motion regularization of

a complete row/column of the images. Although it is useful for traditional non-PGD approaches to introduce a further assumption on the smoothness of the flow field using a pre-filtering process [5], the PGD approach provides in this case a further gain of accuracy and computation time on pre-filtered images, as shown in Table 4. Similarly, while a higher level of information is available on color images (for each pixel, one information stems from each color channel), the use of the PGD-approach remained beneficial at low resolution/SNR even on color images.

Secondly, the gain in terms of computation time is explained as follows: While the non-PGD approach necessitates the solving of a system of $N_1 \times N_2$ equations, the PGD approach relies on the successive resolution of two reduced systems, each one having $2 \times N_1$ and $2 \times N_2$, respectively. In other words: While image dimensions have a multiplicative impact on the equation number of the solved system using the non-PGD approach, an additive impact is substantially obtained using the PGD. Moreover, it is shown in Tables 1 and 2 that the use of noisy data doesn't penalize the computation time when using the PGD, unlike the non-PGD case. The PGD method, which relies on a "global" approach, introduces an inherent image filtering arising from the data reduction process described above. In comparison, the non-PGD algorithm requires a high number of additional iterations in order to cope with the fact that all local noise-related grey level intensity variations are attributed to motion. It is interesting to underline that the OF task is a computationally intensive task that is now usually accelerated using parallel architectures such as Graphical Processing Units (GPU), especially for high resolution images [22]. Such hardware is however intrinsically unable to speed up the calculation for the particular case of low resolution images. Our PGD-based approach is in contrast of particular interest for such situations.

However, it must be underlined that: When the image resolution is increased, the additional time loss arising by the calculation of a new tensorial product using the PGD approach is balanced by the benefit provided by the use of the multi-resolution scheme with the non-PGD case. More specifically, only a few iterations are necessary using the traditional non-PGD approach when the algorithm iterates locally close to the final solution. On the contrary, the PGD-approach necessitates the solving of the two systems of Eq. (2.33) and (2.36) which are, in such a case, more time consuming since they seek a more "global" transformation. A good strategy would be to employ the PGD-based approach for low resolution levels (*i.e.* lower or equal than 128×128), leaving high resolution levels performed by the traditional non-PGD approach. That way, both the accuracy of the motion estimates and the computation time should be improved for all resolution levels, even the highest ones for which the non-PGD would iterate around more accurate values.

We restrained this study on OF methods composed by a robust energy in the data fidelity term and a quadratic penaliser for the regularization term. However, regularization terms with linear growth at infinity, with respect to gradients, play an important role in image processing [10]. Future work will thus concern OF methods composed by a robust energy in both the data fidelity and the regularization term. For such a regularization class, the adequate functional framework is the space of bounded variation functions $BV(\Omega)$. Unfortunately, in such a case, the convergence study detailed in this paper will not be valid anymore: The space $BV(\Omega)$ has a non-tensorial structure and thus classical PGD methods cannot be performed in such situations. This class of regularization will therefore be part of future research. We also believe that the method should be greatly beneficial in terms of computation time

for OF applied on 3D data: Since image dimensions have an additive impact on the number of solved equations using the PGD approach, we anticipate that the benefit in terms of computation time should be even more important than the one presented here in the 2D case. We also believe that it may be possible to extend the proposed PGD approach for other regularization problems such as compressed sensing, image filtering, image segmentation or in-painting (an overview of variational problems in image processing can be found elsewere [28]).

The proposed PGD-based approach finally opens great perspectives for the specific context of real-time applications which necessitate the calculation of motion estimates with a reduced latency, based on low resolution/low SNR images. An example of application scenario can be found in oncology in the domain of real-time image-guided therapies [27]: The possibility to locally deposit thermal energy in a non-invasive way opens a path towards new therapeutic strategies with improved reliability and reduced associated trauma leading to improved efficacy, reduced hospitalisation and costs [13]. For tumor located in the abdomen, an accurate image-based estimation of organ displacement with a low latency is a key requirement. While MRI appears an image modality of choice, the image-registration algorithm often has to cope with low resolution/low SNR images [12]. The proposed PGD-based OF approach thus opens great perspectives in this domain.

5. Conclusion. This paper proposes a new PGD-based OF approach in a classical framework of Sobolev spaces, *i.e.* optical flow methods including a robust energy for the data fidelity term together with a quadratic penalizer for the regularisation term. The theoretical convergence has been demonstrated, an implementation was proposed and a practical evaluation has been conducted on a freely available image collection undergoing various experimental conditions.

The proposed PGD-based approach was found beneficial in terms of both accuracy and computation time for images containing a weak level of information, namely low image resolution and/or low SNR. In particular, we believe that the proposed PGD-based approach opens great perspectives for the specific context of real-time applications which necessitate the calculation of motion estimates with a reduced latency, based on low resolution/low SNR images.

<div align="center">REFERENCES</div>

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, 1985.

[2] A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. *Journal of Non-Newtonian Fluid Mechanics*, 139:153–176, 2006.

[3] A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. part ii: transient simulation using space-time separated representations. *Journal of Non-Newtonian Fluid Mechanics*, 144:98–121, 2007.

[4] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[5] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[6] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proceedings of Fourth International Conference on Computer Vision (ICCV)*, pages 231–236, 1993.

[7] H. Brézis and E. Lieb. A relation between pointwise convergence of functions and convergence of functionals. *Proceedings of the American Mathematical Society*, 36:486–490, 1983.

[8] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV'04)*, volume 3024, pages 25–36, 2004.

[9] É. Cancès, V. Ehrlacher, and T. Lelièver. Convergence of a greedy algorithm for high-dimensional convex nonlinear problems. *Mathematical Models and Methods in Applied Sciences*, 21(12):2433–2467, 2011.

[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[11] F. Chinesta, P. Ladevèze, and E. Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395–404, 2011.

[12] B. Denis de Senneville, A. El Hamidi, and C. T. W. Moonen. A direct PCA-based approach for real-time description of physiological organ deformations. *IEEE Transactions on Medical Imaging*, 34(4):974–982, 2015.

[13] B. Denis de Senneville, M. Ries, L. W. Bartels, and C. T. W. Moonen. MRI-guided high-intensity focused ultrasound sonication of liver and kidney. In *Interventional Magnetic Resonance Imaging Medical Radiology*, pages 349–366, 2012.

[14] A. Falco and A. Nouy. Proper generalized decomposition for nonlinear convex problems in tensor banach spaces. *Numer. Math.*, 121:503–530, 2012.

[15] D. J. Fleet and Y. Weiss. Optical flow estimation. In *Handbook of Mathematical Models in Computer Vision (PDF), Springer*, pages 237–257, 2006.

[16] S. Heldmann, T. Polzin, A. Derksen, and B. Berkels. An image registration framework for sliding motion with piecewise smooth deformations. In *Scale Space and Variational Methods in Computer Vision*, volume 9087, pages 335–347, 2015.

[17] B. K. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[18] P. Ladevèze. On a family of algorithms for structural mechanics. *CR Acad Sci Paris*, 300(2):41–44, 1985.

[19] P. Ladevèze, D. Néron, and P. Gosselet. On a mixed and multiscale domain decomposition method. *Computer Methods In Applied Mechanics and Engineering*, 96:1526–1540, 2007.

[20] P. Ladevèze and A. Nouy. A multiscale computational method with time and space homogenization. *Comptes Rendus Mécanique*, 330(10):683–689, 2002.

[21] P. Ladevèze, J.C. Passieux, and D. Néron. The latin multiscale computational method and the proper generalized decomposition. *Computer Methods In Applied Mechanics and Engineering*, 199(21-22):1287–1296, 2010.

[22] K. Ø. Noe, B. Denis de Senneville, and T. S. Sørensen U. V. Elstrøm, K. Tanderup. Acceleration and validation of optical flow based deformable registration for image-guided radiotherapy. *Acta Oncologica*, 47(7):1286–1293, 2008.

[23] A. Nouy. Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems. *Archives of Computational Methods in Engineering - State of the Art Reviews*, 17:403–434, 2010.

[24] I. Pratikakis, C. Barillot, P. Hellier, and E. Mémin. Robust multiscale deformable registration of 3D ultrasound images. *International Journal of Image and Graphics*, 3(4):547–566, 2003.

[25] M. Schmitt. The (non)tensor structure of bv-spaces. *Results in Mathematics*, 39:91–98, 2001.

[26] D. Sun, S. Roth, and M.J. Black. Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, pages 2432–2439, 2010.

[27] C. Tempany, N. McDannold, K. Hynyen, and F. Jolesz. Focused ultrasound surgery in oncology: overview and principles. *Radiology*, 36(5):39–56, 2011.

[28] L. A. Vese and C. Le Guyader. Variational methods in image processing. In *Chapman and Hall/CRC Mathematical and Computational Imaging Sciences Series*, pages 1–406, 2015.

[29] M. Werlberger, T. Werner, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber-L1 Optical Flow. In *Proceedings of British Machine Vision Conference (BMVC'09)*, 2009.

[30] C. Zachiu, N. Papadakis, M. Ries, C. T. W. Moonen, and B. Denis de Senneville. An improved optical flow tracking technique for real-time mr-guided beam therapies in moving organs. *Physics in Medicine and Biology*, 60(23):9003, 2015.