Robust real time constrained estimation of respiratory motion interventional MRI on mobile organs

Sébastien Roujol, Jenny Benois-Pineau, Baudouin de Senneville, Mario Ries, Bruno Quesson, Chrit T. W. Moonen

Robust real time constrained estimation of respiratory motion for interventional MRI on mobile organs.

Journal:	*Transactions on Information Technology in BioMedicine*
Manuscript ID:	TITB-00283-2011.R1
Manuscript Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Roujol, Sébastien; Laboratory for Molecular and Functional Imaging, FRE 3313 CNRS/Université Bordeaux Benois-Pineau, Jenny; LaBRI, UMR 5800 CNRS/Université Bordeaux 1 Denis de Senneville, Baudouin; Laboratory for Molecular and Functional Imaging, FRE 3313 CNRS/Université Ries, Mario; Laboratory for Molecular and Functional Imaging, FRE 3313 CNRS/Université Quesson, Bruno; Laboratory for Molecular and Functional Imaging, ; Centre de recherche cardio thoracique, Inserm U1045, Université Victor segalen bordeaux 2 Moonen, Chrit; Laboratory for Molecular and Functional Imaging, FRE 3313 CNRS/Université
TIPS:	Image registration, Motion analysis, biomedical image processing, magnetic resonance imaging

Robust real time constrained estimation of respiratory motion for interventional MRI on mobile organs.

Sébastien Roujol*, Jenny Benois-Pineau, *Member, IEEE,* Baudouin Denis de Senneville, Mario Ries, Bruno Quesson, Chrit Moonen

***Abstract*—Real time magnetic resonance imaging is a promising tool for image-guided interventions. For applications such as thermotherapy on moving organs, a precise image-based compensation of motion is required in real time to allow quantitative analysis, retro-control of the interventional device, or determination of the therapy endpoint. Reduced field-of-view imaging represents a promising way to improve spatial and / or temporal resolution. However, it introduces new challenges for target motion estimation since structures near the target may appear transiently due to the respiratory motion and the limited spatial coverage.**

In this paper, a new image based motion estimation method is proposed combining a global motion estimation with a novel optical flow approach extending the initial Horn & Schunck (H&S) method by an additional regularization term. This term integrates the displacement of physiological landmarks into the variational formulation of the optical flow problem. This allowed for a better control of the optical flow in presence of transient structures. The method was compared to the same registration pipeline employing the H&S approach on a synthetic dataset and *in vivo* image sequences. Compared to the H&S approach, a significant improvement ($p<0.05$) of the DICE similarity criterion computed between the reference and the registered organ positions was achieved.

***Index Terms*—Image registration, Motion analysis, Biomedical image processing, Magnetic resonance imaging.**

I. Introduction

REAL time magnetic resonance (MR) imaging proved to be a promising candidate for guiding non- and mini-invasive surgical interventions [1]. The additional use of quantitative MR-measurements of tissue specific properties such as MR-relaxation times or the local temperature allows monitoring of the therapeutic progress and helps to determine the endpoint of the intervention. However, quantitative measurements require motion correction to enable a direct comparison on a voxel-by-voxel basis between the dynamically acquired images. In the particular case of MR-guided thermal ablations, real time MR-thermometry is used to compute temperature maps inside the human body [2]. The temperature evolution can be used as an accurate and immediate prediction of tissue necrosis [3] and thus requires all temperature maps registered to a common position. Moreover, the whole registration process has to be realized within the interval of two MR-acquisitions to ensure the real time condition.

Typically, interventional procedures are restricted to a part of the organ/tissue under study making reduced field of view (FOV) imaging desirable. This would allow improvement of the spatial and / or temporal resolution in order to decrease partial volume effects (undesirable for quantitative analysis) and to increase the imaging framerate (required to observe rapid phenomena). Several strategies have been proposed that use saturation slabs [4], outer volume suppression [5] or interactive reduced FOV imaging [6]. However, a reduced FOV may introduce new challenges for the target motion estimation since structures may appear transiently due to the respiratory motion and the limited spatial coverage.

A variety of motion estimation algorithms [7], [8] have been suggested in the field of medical imaging. Optical flow algorithms [9] have been proposed for motion estimation on abdominal organs for MR-guided laser ablation [10]. Recently, an approach using an initial global motion estimation followed by an optical flow algorithm was developed for real time MR-thermometry in abdominal organs [11]. Optical flow algorithms allow the estimation of a velocity field assuming an intensity conservation during displacement, mathematically expressed by the optical flow equation (OFE):

$$I_x u + I_y v = -I_t. \quad (1)$$

Here u and v are the displacement vector components, and $I_{x,y,t}$ are the spatio-temporal partial derivatives of the image pixel intensity. However, a direct estimation by minimizing the deviation from OFE (equation 1) is an under-determined problem and an additional constraint is required. The method proposed by Horn and Schunck (referred to as H&S in this paper) introduces additional physical constraints enforcing the smoothness of the motion field [12]. They seek $u(x, y)$ and

This work was supported in part by Ligue Nationale Contre le Cancer, Conseil Régional d'Aquitaine, Diagnostic Molecular Imaging, Agence National de Recherche, Fondation InNaBioSanté and Philips Medical System.

M. Roujol*, Denis de Senneville, Ries, Quesson and Moonen are with the Laboratory for Molecular and Functional Imaging: from Physiology to Therapy FRE 3313 CNRS/ University of Bordeaux 2 - 146 rue Leo Saignat, Case 117, 33076 Bordeaux, France. e-mail: {s.roujol,m.ries,b.quesson,chrit.moonen,baudouin}@imf.u-bordeaux2.fr. M. Roujol and Mrs Benois-Pineau are with Laboratoire Bordelais de Recherche en Informatique, UMR 5800 CNRS/University of Bordeaux 1 - 351, cours de la Liberation, 33405 Talence, France e-mail: jenny.benois@labri.fr. M. Quesson is also with Centre de recherche cardio thoracique, Inserm U1045, University of Bordeaux 2 - 146 rue Leo Saignat, 33076 Bordeaux, France. *Asterisk indicates the corresponding author.*

$v(x, y)$ minimizing:

$$E(u, v) = \iint \left(\left[I_x u + I_y v + I_t \right]^2 + \alpha^2 \left[\|\nabla u\|_2^2 + \|\nabla v\|_2^2 \right] \right) dxdy, \quad (2)$$

where α^2 is a weighting factor designed to link both intensity variation and motion field regularity and ∇u and ∇v are the spatial gradients of $u(x, y)$ and $v(x, y)$, respectively.

Using reduced FOV imaging, the assumption of energy conservation may be violated due to the potential presence of transient structures. In such conditions, Horn & Schunck's algorithm is expected to fail locally (near the transient structure) to recover the correct motion. This is known as one of the ill-posed problems in motion estimation, namely the occlusion problem [12]. In detail, if a structure is only present in the reference image, the algorithm might try to match the signal from this structure with a different structure in the image to be registered. To improve the robustness of the algorithm against this effect, one can increase the value of α^2 to increase the smoothness constraint of the motion field. However, this will reduce the ability of the algorithm to provide reliable motion estimation in presence of complex deformations.

Therefore, the first step of this study was to hypothesize that a combination of intensity and landmark registration could be used to improve the robustness of the registration against transient structures. Such combinations have been proposed in the past such as in [13] where a large number of landmarks were employed for brain registration. This method was further extended to deal with a smaller number of points (manually defined) for interactive registration of medical images [14]. In these approaches, both intensity and landmark metrics were minimized sequentially inside an iterative minimization process. Several unified minimization frameworks were also proposed. In [15] the algorithm was designed to obtain a final motion field fitting each pre-estimated landmark motion. In the presented application, due to the low SNR, a non-negligible uncertainty of the landmark correspondence may be expected and would thus strongly affect such registration approach. Another solution proposed by Becciu *et al.* [16], attempts to use tags obtained from an MR-tagged sequence in their registration method for cardiac contraction assessment. Unfortunately, this MR-sequence is generally unsuitable for interventional MRI, since images are tagged by regular lines where the signal has been removed. Recently, a variational approach, integrating segmented region motion, was proposed for large displacement estimation [17]. This method uses a linearized OFE deviation together with regularization terms which include the correspondence of region displacements in the image plane. Despite the interest of such an approach in general purpose video sequences, its application to MRI sequences is not straightforward due to the inherent difficulties of segmentation of frames into spatially coherent regions. Recently, a grid-based deformation model was proposed [18], but was also shown to be sensitive/limited by the landmark extraction process according to the authors.

In this paper, we propose a new real-time motion estimation method for MRI sequences which can operate in the case of reduced FOV imaging. The preliminary results of this approach can be found in [19]. The contribution of the paper is twofold. We first define the two stage interventional protocol for a robust choice of constraint points and then formulate the constrained optical flow estimation by introducing an additional regularization term in the H&S method. In our formulation, the introduction of a smooth weighting function allows for a local control of the influence of constraint points. Furthermore, in order to ensure the real-time requirement together with a short latency, all computationally intensive calculations are off-loaded to a dedicated graphics processing unit (GPU). The proposed algorithm is referred to as constrained motion estimation (CME) in the scope of this paper. It was compared with the same registration pipeline employing the H&S optical flow approach. Algorithm evaluations were conducted on both synthetic data and cardiac and kidney MR-images of healthy volunteers under free breathing conditions.

II. Material and Methods

The proposed CME algorithm is a two-step procedure (see Fig. 1). The first step consists of selecting the constraint points along the boundary of the organ in the reference image of the time series. In a second step the motion is estimated for each image as follows: a global translational motion estimation is performed and used to initialize a local estimation of the displacement of constraint points. Non-physiological constraint point displacements are automatically identified and corresponding constraint points are discarded. The displacements of the constraint points are then integrated into the constrained optical flow algorithm (using the global estimated motion as preconditioning) to obtain the final motion field. A detailed description of each algorithm step is presented in the next sections.

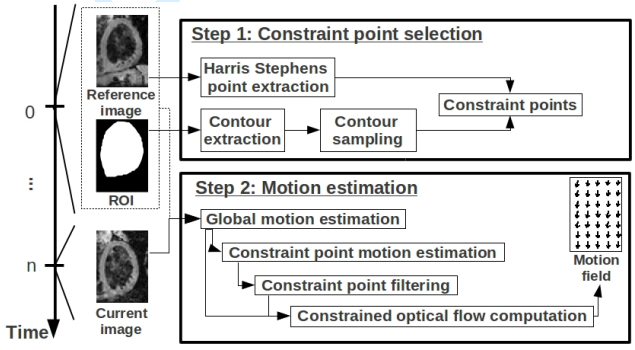

Fig. 1: General scheme of the algorithm. Prior to the intervention, constraint points are automatically extracted from a reference image (Step 1). Then, during the procedure, the motion field is estimated for each frame (Step 2).

A. Step 1: Constraint point selection

Anatomical points are localized and tracked over time in order to guide and to constrain the motion estimation of the target. To select them, anatomical structures such as organ boundaries, which remain present during the acquisition and

follow the target, are suitable. For this, a region of interest (ROI) is manually set around the target of the reference image. The contour of the ROI is first extracted and then regularly sampled in space to obtain a set of N points surrounding the target. To allow a certain degree of freedom on the ROI drawing, a refinement of the location of the N points is performed by moving them near feature points computed on the reference image. Due to the abundance of works in stereo matching and image retrieval, a large amount of feature point detectors has been tested and reported in literature [20]. The critical point is the stability of these methods with respect to affine transformations of the image plane, lightening, scale variations and noise. In the case of MR images, the noise and deformable motion are the main factors. According to the evaluation in [20], the Harris-Stephens detector [21] appeared to provide a good compromise between robustness and computation time. The feature point detection is based on the following response function:

$$R(x,y) = Det(M_{x,y}) - k.Tr(M_{x,y})^2, \tag{3}$$

with

$$M_{x,y} = \sum_{i,j \in S} w_{i,j} \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix}_{i,j}, \tag{4}$$

where (x,y) denotes the spatial coordinates, Det denotes the determinant of a matrix, Tr is the trace of a matrix, w is a weighting factor (Gaussian kernel over the region S centered on (x,y)) and k is a sensitivity parameter. The value of k is generally chosen in the [0.04-0.15] interval. A value of 0.04 was determined empirically suitable for our application. The response is positive in a corner region. Therefore, the feature point with the higher response in a small neighborhood (e.g. 3x3 pixels) of the contour point is selected as a landmark point. The small neighborhood search was designed to prevent positioning a landmark on a different structure/organ that may depict a different motion from the targeted organ. If no feature point is present, the initial ROI contour point is then selected.

B. Step 2: motion estimation algorithm

An optical flow based algorithm is more efficient when it is initialized after compensating for large displacements. Therefore, to initialize it, a global motion estimation is first performed with a simple translational model. The translation parameters (horizontal and vertical) are estimated using a sign-gradient-descent with fixed step inspired by the Netravali-Robbins method [22]. The estimation is restricted to the ROI defined in section II-A.

Constraint point displacements are then individually estimated (two translation parameters) using the global estimated displacement as the initial estimate. This estimation is restricted to a small patch centered on each constraint point intersected with the initial ROI to allow a local refinement of the global displacement. We experimentally found that a patch size of 10×10 pixels was satisfactory using a given voxel size of 2-3 mm obtained with the employed MR-acquisitions.

To remove occasional non-physiological estimates, the following outliers rejection was added. The displacement vector (d_x, d_y) of a constraint point was assumed to follow a bivariate Gaussian distribution with independence of d_x and d_y coordinates. A constraint point was automatically rejected if at least one of its displacement components violated the marginal 3-sigma rule.

The idea of the presented approach is to constrain the H&S formulation by locally estimated displacements of feature points. Hence, we propose the following extension of the Horn & Schunck formulation with an additional regularization term:

$$E_c(u,v) = \iint \left(\left[I_x u + I_y v + I_t \right]^2 + \alpha^2 \left[\|\nabla u\|_2^2 + \|\nabla v\|_2^2 \right] + \lambda^2 \sum_{i=1}^{N} \left(\rho(d_i, R) \left[(u-u_i)^2 + (v-v_i)^2 \right] \right) \right) dxdy, \tag{5}$$

where (u_i, v_i) are the horizontal and vertical components of the displacement estimated for the i^{th} constraint point. λ^2 is the regularization parameter that allows balancing between the initial behavior of the H&S algorithm and the constraint influence. ρ is a distance function, defined as

$$\rho(d,R) = \exp\left(-d^2/R^2\right) \tag{6}$$

where d represents the Euclidean distance between the pixel of coordinates (x,y) and the i^{th} constraint point, R is a bandwidth parameter. To minimize $E_c(u,v)$, we used the calculus of variation and obtained the following system:

$$\begin{cases} I_x^2 u + I_x I_y v = \alpha^2 \nabla^2 u - I_x I_t + \lambda^2 \sum_{i=1}^{N} \left(\rho(d_i, R) u_i \right) \\ I_x I_y u + I_y^2 v = \alpha^2 \nabla^2 v - I_y I_t + \lambda^2 \sum_{i=1}^{N} \left(\rho(d_i, R) v_i \right) \end{cases}. \tag{7}$$

Then, we used the approximation of the Laplacian, as suggested by Horn and Schunck, where $\nabla^2 u = \overline{u} - u$, with $\overline{u}$ the mean value of u in the neighborhood (3x3 pixels) of the estimated point [12]. Therefore, the system can be rewritten as:

$$\begin{cases} a_{11} u + a_{12} v = b_1 \\ a_{21} u + a_{22} v = b_2 \end{cases}, \tag{8}$$

with

$$\begin{cases} a_{11} = I_x^2 + \alpha^2 + \lambda^2 \sum_{i=1}^{N} \rho(d_i, R) \\ a_{12} = a_{21} = I_x I_y \\ a_{22} = I_y^2 + \alpha^2 + \lambda^2 \sum_{i=1}^{N} \rho(d_i, R) \\ b_1 = \alpha^2 \overline{u} - I_x I_t + \lambda^2 \sum_{i=1}^{N} \rho(d_i, R) u_i \\ b_2 = \alpha^2 \overline{v} - I_y I_t + \lambda^2 \sum_{i=1}^{N} \rho(d_i, R) v_i \end{cases}. \tag{9}$$

Finally, based on the Jacobi method, the system can be solved with the following iterative scheme:

$$u^{n+1} = \frac{b_1^n a_{22} - a_{12} b_2^n}{a_{11} a_{22} - a_{12} a_{21}}, \; v^{n+1} = \frac{a_{11} b_2^n - a_{21} b_1^n}{a_{11} a_{22} - a_{12} a_{21}}. \tag{10}$$

C. Implementation

Registration to a reference frame has been preferred to a concatenation of frame to frame estimations since the latter generally leads to the accumulation of errors, especially when a high framerate is applied such as in abdominal imaging (10-15 Hz). In cardiac applications, since the imaging framerate is generally limited to the cardiac frequency, the respiratory motion between two successive images generally depicts a large amplitude, devaluing the benefit of a frame to frame motion estimation. For both the H&S and the CME implementation, the multi-resolution approach from [23] was used, which refines the motion estimation algorithm from the 3^{rd} level of sub-resolution to the full resolution. We off-loaded the most time consuming task, i.e. the iterative numerical scheme of the optical flow, to a dedicated GPU.

In-vivo calibration of the employed algorithms is complex since it depends on the criterion to be optimized. The calibration of the H&S method (α^2 value) has to deal with contradictory effects. A small regularization of the motion field is required to enable the estimation of complex motion and to have a globally reliable estimated motion. On the other hand, a high constraint on the motion field smoothness would reduce the registration artifact induced by intrusive structures but at the same time will limit the ability of the algorithm to estimate complex motion. This may thus deteriorate the registration in the entire organ. Therefore, we calibrated the employed algorithms in a way to maintain optimal performance in the general case of full FOV imaging (without the presence of intrusive structures). As recently shown [24], a reliable in-vivo calibration of the H&S algorithm was obtained for a range of α^2 values between 0.1-0.5. A plateau was generally observed for these ranges of values and its lower bound was suggested as a good way to cope with variations of the breathing pattern (such as an amplitude variation or drift of the respiration pattern). Therefore, a α^2 value of 0.1 was employed for the H&S algorithm. We empirically found a near-optimal solution for the CME calibration by employing the following parameters: α^2=0.1, λ^2=0.1, N=20 and R^2=5. Note that a similar optimal configuration was obtained for the synthetic dataset experiment (see Result section) except for the α^2 values that were higher due the lowest complexity of the synthesized motion.

The overall algorithm was implemented in C++ and evaluated on a dual processor (INTEL 3.1 GHz Penryn, two cores). The GPU implementation was based on the Compute Unified Device Architecture (CUDA) framework [25] using a NVIDIA GTX280 card.

D. Experimental setup

The proposed algorithm was evaluated on both synthetic and *in vivo* datasets:

1) Synthetic dataset experiment:

Data creation: A sequence of T(= 30) images was created. To simulate respiratory motion typically encountered on mobile organs, a periodic (period=6 frames) geometric transformation composed by a 2D translation (T_x,T_y) and scaling (S_x,S_y) was synthesized ($T_x = \{0, 0.5, 1, 1.5, 1, 0.5\}$, $T_y = \{0, 2.5, 5, 7.5, 5, 2.5\}$ pixels, $S_x = S_y = \{1, 1.03, 1.06, 1.09, 1.06, 1.03\}$). A signal-to-noise ratio (SNR_{dB}) of 1.3 was chosen to simulate a realistic acquisition (typically between 0.7 and 1.3). A structure appearing transiently in the lower part of the image was added in half of the images to simulate the effect potentially encountered with reduced FOV imaging.

Quality assessment of the motion estimation: Since the real motion ($D_{gt} = (u_{GT}, v_{GT})$) and the estimated motion ($D = (u, v)$) are available for each pixel in such synthetic dataset experiment, the measures commonly reported in the optical flow community such as the endpoint error (EE) and the angular error (AE) of the flow [9], [26] were computed, with:

$$EE = \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2}, \tag{11}$$

$$AE = \cos^{-1} \frac{1 + u \times u_{GT} + v \times v_{GT}}{\sqrt{1 + u^2 + v^2}\sqrt{1 + u_{GT}^2 + v_{GT}^2}}. \tag{12}$$

In addition, to provide additional information about the smoothness of the estimated motion field, the harmonic energy of the estimated flow [27] was computed.

2) In vivo experiments:

In vivo experiments were conducted on the heart and the kidney of a total of 12 healthy volunteers. The MRI scans were performed under free breathing conditions using a 1.5 Tesla scanner (Philips Achieva/Intera, Best, The Netherlands). The volunteers were positioned head first in the supine position. The MR sequences employed saturation slabs to obtain structures appearing transiently in the FOV as typically encountered when zoom imaging is used (these conditions are further referred to as "zoom imaging conditions").

***In-vivo* study on the heart under zoom imaging conditions:** Dynamic MRI was performed on the hearts of six healthy volunteers. The acquisition sequence was ECG-gated (to observe the heart in the same cardiac phase) using a five element phased array cardiac coil. Five contiguous adjacent slices were acquired per cycle (200 cycles per scan), in short axis view, at end diastolic phase. A slice tracking technique [28] was used to compensate for respiratory motion in the third dimension. Blood signal reduction was obtained using saturation slabs positioned on each side of the imaging stack. The single shot EPI sequence employed the following parameters: FOV=260 × 260 mm^2, voxel size=2.7 × 2.7 × 7 mm^3, echo time=20 ms, repetition time=40 ms, SENSE acceleration factor=1.6 [29]. A saturation slab was positioned underneath the extreme position of the heart (corresponding to the position at maximum respiratory displacement) to simulate zoom imaging conditions.

***In-vivo* study on the kidney under zoom imaging conditions:** 200 frames (single slice) in coronal orientation were acquired using a four element phased array body coil. A dual shot EPI sequence employed the following parameters: FOV=200 × 400 mm^2, voxel size=2.3 × 2.3 × 6 mm^3, echo time=26 ms, repetition time=52 ms, flip angle=35°. Zoom imaging conditions were achieved using a saturation slab positioned on the top of the extreme position of the kidney.

Quality assessment of the motion estimation: In such conditions, the typical amplitude of both heart and kidney

motion is about 8 pixels in the imaging plane between two extreme images in the respiratory cycle. Since the true motion is unknown the quality assessment of the registration was analyzed by computing the DICE similarity coefficient (DSC) [30] between the position of the organ (ROI_t) in each registered frame (t) and its position in the reference frame (ROI_{ref}), as follows:

$$DSC(t) = \frac{2(ROI_{ref} \cap ROI_t)}{ROI_{ref} + ROI_t}, \quad (13)$$

Each ROI was obtained by manual segmentation. A DSC value of 1 indicates an ideal registration (perfect ROI matching). The harmonic energy has also been computed and reported for all tested cases to assess the energy of the deformation fields.

III. Results

A. Synthetic dataset experiment

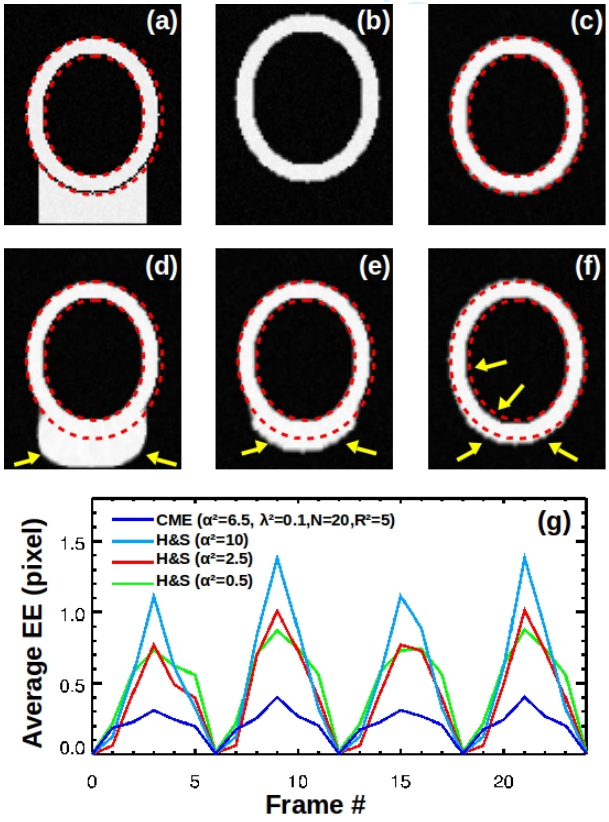

Fig. 2: Registration example obtained on the synthetic dataset experiment. The reference image (a) and the image to be registered (b) are displayed together with the registered images obtained with the CME (c) and the H&S approach using a value of α^2 of 0.5 (d), 2 (e) and 6.5 (f). The time evolution of the average endpoint error (EE) inside the target (between both dashed red ellipses) is reported in (g).

An example of synthetic images is shown in Fig. 2 where the reference image (Fig. 2a, containing the transient structure) and an image corresponding to the maximum synthesized motion (Fig. 2b, 4^{th} image of the cycle) are displayed. The contours (interior and exterior) of the target obtained from the reference image are reported in dashed red curves. The H&S approach was employed to register the image in Fig. 2b using different α^2 values. When using a small value ($\alpha^2 = 0.5$), the registration is strongly influenced by the underlying structure leading to a severe registration artifact (yellow arrows in Fig. 2d). By using an optimally calibrated value ($\alpha^2 = 2.5$, see Fig. 3a), the registration artifact was reduced but still present (yellow arrows in 2e). A higher value of $\alpha^2 (= 10)$ is reported in Fig. 2f where the registration artifact introduced by the transient structure appeared substantially reduced due to a higher weight on the smoothness of the motion field. However, in this case a different registration artifact was observed due to the inability of the method to handle complex motion (here the scaling effect) as shown by the yellow arrows. The CME approach provided a reliable registration where the registered target perfectly matched the reference target contour. These results were confirmed in the plot of the averaged EE values over time (computed inside the target, between the two red dashed curves) in Fig. 2g. While the H&S approach periodically failed to recover the correct motion, the CME offered more robust performances over time.

In order to better characterize the potential gain and limitations of the compared methods, their performance in terms of averaged error endpoint, harmonic energy and averaged angular error were precisely investigated as a function of the employed parameters (Fig. 3). As previously observed in Fig. 2d, low α^2 values (Fig. 3a, 3e and 3i) provided poor performance since the methods became very sensitive to the presence of the intrusive structure. On the contrary, high values limited the ability of the algorithm to estimate complex motion and also deteriorated the motion estimates as confirmed by the convergence of the harmonic energy toward a very small value. The averaged EE values obtained with an optimal α^2 calibration were 0.42 for the H&S approach ($\alpha^2 = 2.5$) and 0.21 for the CME method (α^2=6.5) showing a reduction of the averaged EE by a factor of 2. The λ^2 value influence was then investigated ((Fig. 3b, 3f and 3j). As expected, small λ^2 values tend to the H&S performance and high values tend to the extrapolation of the constraint point motions (and their associated uncertainty) leading in both cases to a deterioration of the registration. A good calibration of λ^2 (=0.1, blue curve) provided a significant improvement of the motion estimation (Fig. 3b, 3j). The influence of the constraint point number (N) was then evaluated (Fig. 3c, 3g and 3k). Although an optimal value was reached around 20 points (Fig. 3c, 3k, blue curve), the sensitivity of the CME to this parameter was limited. Finally the influence of the bandwidth (R^2) provided an optimal calibration for a value of 5 (Fig. 3d, 3h, 3i, blue curve).

B. In vivo experiments

Similar results were obtained in *in vivo* experiments in both the heart and the kidney of healthy volunteers. Registration examples are shown for both organs in Fig. 4. By comparing the reference images (Fig. 4a, 4i) with images acquired at different positions in the respiratory cycle (Fig. 4b, 4j), one can observe some structures appearing transiently (see yellow arrows) due to the signal cancellation obtained from the

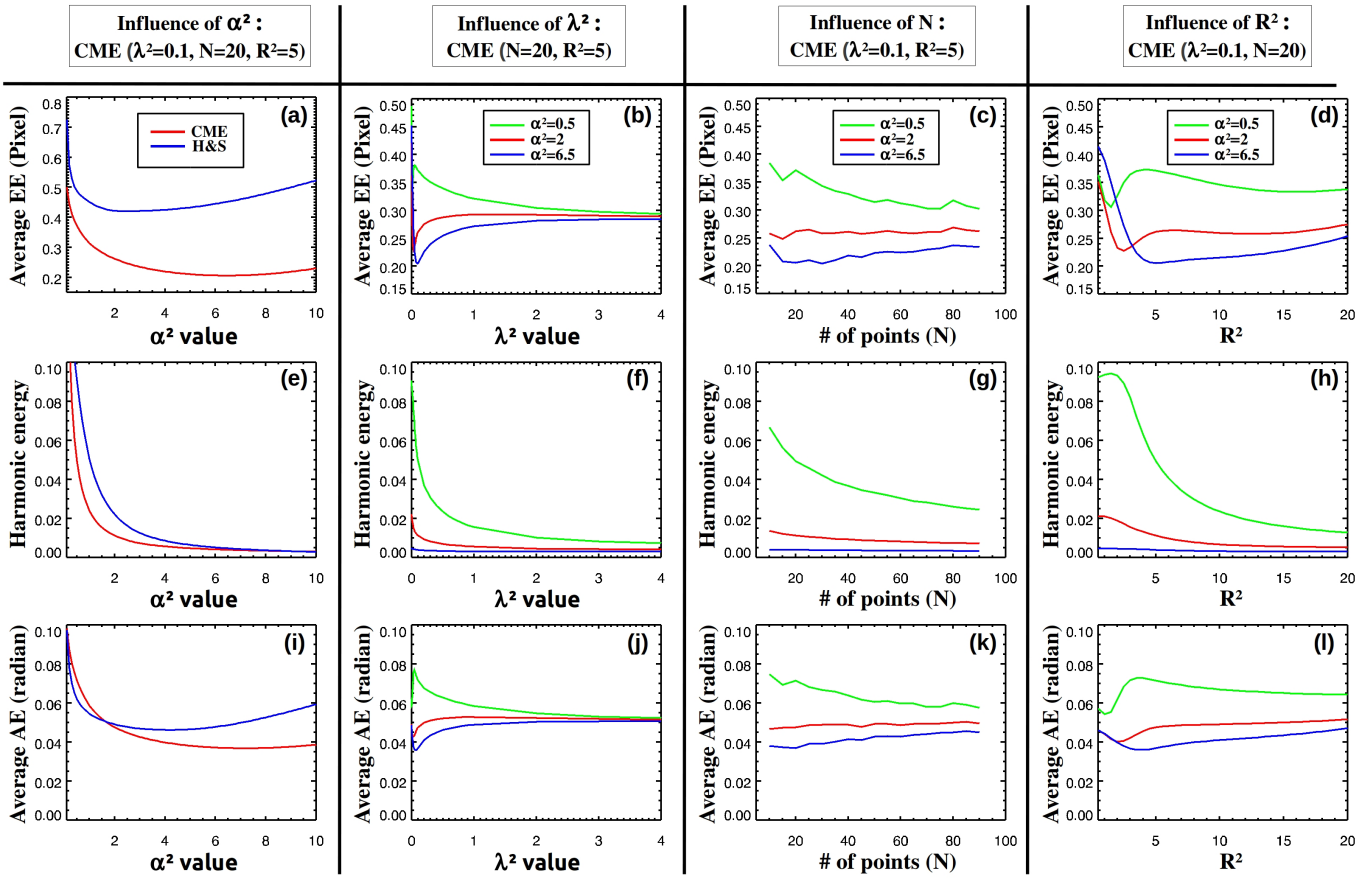

Fig. 3: Influence of the free parameter tuning on the H&S and the CME methods in the synthetic datasets. The averaged error endpoint (EE), the harmonic energy (HE) and the average angular error (AE) are reported in the first, second and third lines of the figure respectively. Note that the averaged EE and AE computation were restricted to the target area (defined between the two red dashed ellipses in Fig. 2a). The influence of the motion field weighting (α^2) on both the H&S and the CME approaches are displayed in the first column. Then, other CME parameters were analyzed: the constraint point term weighting (λ^2, second column), the number of constraint points (N, third column) and the ray influence of the constraint point (R^2, forth column).

saturation bands. In the cardiac images, the area with low signal intensity in the lower part corresponds to a saturation band, the latter may allow for the reduction of the FOV without additional fold-over artifacts. The signal of the liver (below the heart) almost disappeared due to the displacement induced by the respiration. In the abdominal images, two perturbations were observed in the top part of the kidney: the liver, above the kidney, partially disappeared in Fig. 4j and the intensity of the upper part of the kidney depicted a high variation. In both examples, while the registration obtained with the H&S approach (Fig. 4c, 4k) was severely deteriorated in the regions near the transient structures (see red arrows), a reliable registration was obtained in the totality of the organs using the CME (Fig. 4d, 4l).

These findings were typical for the entire sequences as confirmed by the time evolutions of the DICE similarity coefficient and the harmonic energy, respectively shown in Fig. 4e, 4m and Fig. 4f, 4n. Due to the respiratory cycle, the transient structures appeared periodically in the time series and the H&S approach periodically failed to recover a reliable motion estimate leading to low DICE similarity coefficient and elevated harmonic energy values. The CME clearly outperformed the H&S approach by providing a better overall registration and more stable performance.

Over the 12 volunteers, the averaged DICE similarity coefficient (central point inside the box) obtained with the H&S method has been significantly improved ($p<0.05$) using the proposed CME as shown in Fig. 4g and 4o. The minimal DICE similarity coefficient values were typically very low for certain frames using the H&S method, whereas the proposed CME allowed maintaining a better performance for all the frames (around 0.92 and 0.96 for the heart dataset and the kidney dataset, respectively). As expected, higher harmonic energy values were obtained with the H&S method since the method periodically failed and employed a relative lower weight of the smoothness constraint term (the same α^2 value (=0.1) was employed for both algorithms).

Constraint point filtering allowed the rejection of constraint points with non-physiological estimated displacement. Averaged over all volunteers, less than 0.37 % and 2.67 % of the constraint points were rejected for the heart and the kidney dataset registration, respectively (with a maximum of two

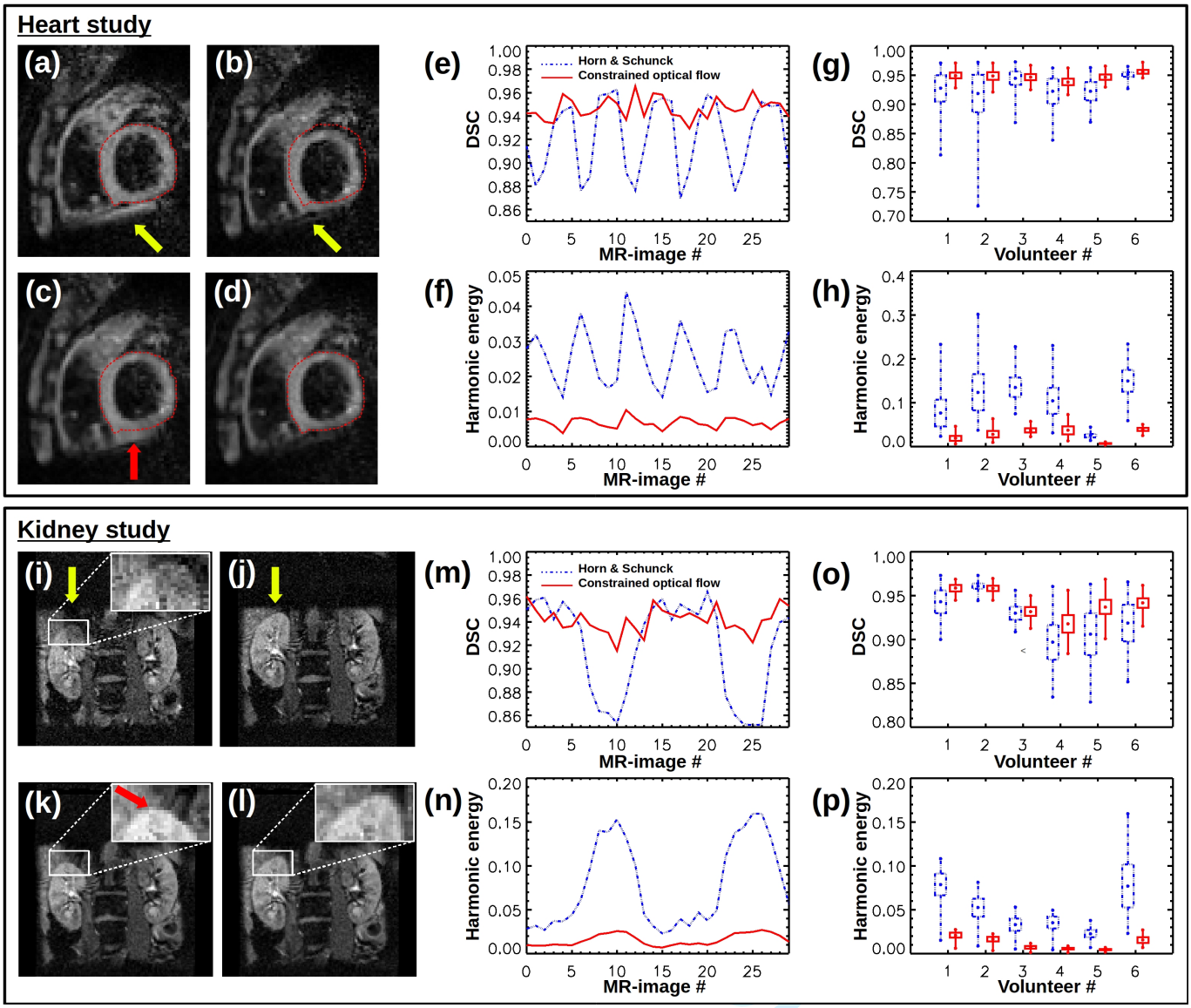

Fig. 4: Registration result obtained *in vivo* in the heart and the kidney of free breathing volunteers. Registration examples in the heart (volunteer #5) and in the kidney (volunteer #6) are given where the reference images (a,i), the images to be registered (b,j) and the registered image obtained with the H&S (c,k) and the CME (d,l) methods are displayed. The time evolution of both the DICE similarity coefficient (DSC) and the harmonic energy (HE) are shown for each registration example in (e,m) and (f,n), respectively. Finally, the DSC (g,o) and HE (h,p) are reported for each volunteer over time as Box and Whisker plot where the minimum (lower point), maximum (upper point), average (point inside the box) and standard deviation (box height) values are shown. While the H&S method was disturbed by the presence of transient structures, the CME provided robust and reliable registration performance for all the frames.

constraint points per frame.

C. Real time benchmarking

Benchmarking was realized for each processing step for an image sequence of spatial resolution 128 × 128. The GPU based implementation allowed a significant reduction of the whole computation time. The total computation time of the proposed approach was evaluated to 22 ms (against 87.5 ms using a CPU only implementation) and was composed by: global motion estimation (5 ms), motion estimation of constraint points (10 ms), optical flow iterative scheme (7 ms) and image registration (0.01 ms). An acceleration factor of 10 was achieved for the computation time of the iterative numerical scheme of the optical flow algorithm (equation 10).

IV. Discussion

A. Performance of the proposed method

On the synthetic dataset, the proposed CME outperformed the H&S approach that failed to estimate the real motion in presence of structures appearing transiently. The endpoint error of flow obtained with the optimal H&S calibration has been

reduced by a factor of 2 using the proposed CME as shown in Fig. 3a.

Furthermore, the volunteer studies confirmed the *in vivo* feasibility of the CME in both the heart and the kidney. A reliable registration was obtained in the totality of the organ in all frames. Using the H&S method, similar performance was also achieved in frames with similar structures in the reference image. However, a severe degradation of its performance was observed in the presence of additional intrusive structures. Although this artifact could have been reduced by increasing the regularization of the motion field, this would have, at the same time, decreased its ability to estimate complex motion (as can be observed in Fig. 2f and 3a using a high value of α^2). As recently shown [24], such effect is generally observed with the H&S method by using α^2 values above the interval [0.1-0.5] with the employed *in-vivo* images.

B. Real time feasibility of the method

MR-guidance of interventional procedures relies on the instantaneous availability of the processed images. Therefore, this limits the available computation time. In addition, Denis de Senneville *et al.* demonstrated in [31] that large latencies have to be compensated with the help of accurate motion prediction. However, the performance of the prediction algorithm increases greatly with short latencies. Recently, in the particular case of HIFU ablation on mobile organs, it was demonstrated that a latency smaller than 100 ms was required for the adjustment of the beam position in order to ensure an energy deposition similar to a static experiment [32]. Here, the GPU implementation offered an acceleration factor of 10 for the computation of the constrained optical flow which is in accordance with the published work [11], [33]. A significantly higher acceleration factor would be expected using higher resolution as shown in [33] since it would benefit from a larger amount of data to process (higher occupations of each processor) and a lower relative overhead. Overall, the demonstrated CPU/GPU implementation allows the acceleration of the required processing time by a factor of four and thus ensures the real time conditions with a short latency. Further reduction of the latency may be obtained by investigating the use of more complex optimization schemes aiming to improve the convergence speed of the algorithm.

C. Calibration of the CME

Contrary to the previous works using constraint points, a comprehensive formulation of the minimization problem was proposed. In the proposed approach, the confidence into the predetermined displacement of the selected constraint points can be freely adjusted with the regularization parameter λ^2. The quality of the obtained optical flow depends on the quality of the initial constraint point vectors, the number of constraints (N) and the bandwidth (R) of the ρ function parameters:

- For the motion estimation of the constraint points, only a translational model was considered as it was the most robust for small patch sizes surrounding constraint points. The optimal patch size in 128×128 MRI sequences was found to be 10×10 for our images.
- In order to control optical flow, the constraint points have to be placed near eventual occlusion (or problematic area). The manual choice of constraint points is not realistic during an interventional procedure, and we can only encourage the staff physician to approximately trace the contour of the ROI. Hence, the subsampling has to be sufficient in order to get a good coverage of the problematic area. On the other hand, a too large number of constraint points will slow down the computational process and may degrade the registration performance by leading to a quasi interpolation of the constraint displacements and their associated uncertainties. Therefore, for the demonstrated application, N=20 was found to be near-optimal.
- The bandwidth R of the ρ function regulates the influence of remote points. The large bandwidth yields a quasi interpolation of constraint point displacements over the whole image. An optimal experimental value was R=$\sqrt{5}$.
- Outlier rejection for constraint point vectors was found particularly useful for small patch sizes where the estimation is more sensitive to out-of-plane motion, noise, etc.

D. Limitations, clinical perspectives and future works

Due to technical limitations of fast MR acquisition sequences, extensive 3D volume imaging on mobile organs is hard to achieve. The proposed technique has thus been evaluated in the 2D case. An extension of the method towards 3D motion correction may also be considered in future works and may improve the motion correction in the third dimension by reducing out-of-plane motion artifacts. However, although the proposed algorithm can be easily extended to 3D, the main challenge would likely remain in the design of a reliable 3D MR-sequence. In such acquisitions, the scan time increase would render the sequence more sensitive to intra-scan motion and may alter the fat and the blood signal cancellation. Echo volumar imaging [34] may help to decrease the scan time of such 3D acquisition; however, its associated low resolution and its robustness against the latter artifacts should be carefully investigated.

A robust formulation of equation 5 using robust estimators such as the Humber function or Lagrangian [35], especially in the first two terms of equation 5 should be investigated in order to account for both brightness variation and smoothness violation (motion discontinuities). Also, the integration of a median filter at each iteration step of the iterative minimization scheme has to be carefully investigated since significant improvements have been achieved in this direction [36]. Nevertheless, the real time compatibility of such approach has to be evaluated. In addition, the integration of additional regularization terms such as in [17] should be investigated. Although these approaches may improve the motion estimation quality, the balance between robustness, simplicity (with regard to the number of free parameters) and performance has to be carefully investigated in the perspective of clinical use. The presented framework also opens great perspectives for integration of other motion information such as navigator or ultrasonic echoes.

The method shows a promising potential for clinical integration for two reasons. First, from its simplicity it only requires a small intervention by the staff physician (only for the mask drawing that only requires a few seconds (∼5s)) and few free parameters to be tuned *a priori*. Then, the employed values of these parameters were always identical for both heart and kidney studies demonstrating the non-necessity of a re-calibration of the parameters for each sequence and the robustness against the choice of the parameters.

Finally, although the feasibility of the method has been shown on healthy volunteers, its feasibility on other organs such as the liver or on patients with, for example, irregular cardiac motion remains to be investigated. In addition, the method has been tested in conditions of a non invasive procedure, its feasibility in the presence of invasive devices, such as catheters, will have to be evaluated in future studies.

V. Conclusion

In this paper a new regularization constraint of the energy functional of the H&S method was presented. This approach represents a flexible solution to integrate constraint point displacements into the optical flow estimation. This extension has been demonstrated to render optical flow methods well suited to accurately estimate the motion for interventional MRI on mobile organs in presence of intrusive structures in reduced FOV. Significant improvements were achieved compared to the Horn & Schunck approach. Finally, the use of parallel processing on affordable commodity graphics hardware demonstrates the feasibility of the algorithm in real time with very short latency required for interventional procedures.

References

[1] R. Edelman, J. Hesselink, M. Zlatkin, and J. Crues, *Clinical magnetic resonance imaging*. WB Saunders Philadelphia, 1996, vol. 2.

[2] V. Rieke and K. Butts Pauly, "MR thermometry," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 2, pp. 376–390, 2008.

[3] S. Sapareto and W. Dewey, "Thermal dose determination in cancer therapy," *Int. J. Rad. Onc. Biol. Phys.*, vol. 10, pp. 787–800, 1984.

[4] A. Haase, "Localization of unaffected spins in NMR imaging and spectroscopy (LOCUS spectroscopy)," *Magn Res Med*, vol. 3, no. 6, pp. 963–969, 1986.

[5] B. Wilm, J. Svensson, A. Henning, K. Pruessmann, P. Boesiger, and S. Kollias, "Reduced field-of-view MRI using outer volume suppression for spinal cord diffusion imaging," *Magn Res Med*, vol. 57, no. 3, pp. 625–630, 2007.

[6] T. Schaeffter, V. Raschea, P. Bornerta, and G. Mensb, "Interactive reduced FOV imaging for projection reconstruction and spiral acquisition," *Magn Res Imaging*, vol. 19, no. 5, pp. 677–684, 2001.

[7] J. Maintz and M. Viergever, "A survey of medical image registration," *Med Image Anal*, vol. 2, no. 1, pp. 1–36, 1998.

[8] T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q. Pham, T. Katila, and I. Magnin, "A review of cardiac image registration methods," *IEEE Trans Med Imaging*, vol. 21, no. 9, pp. 1011–21, 2002.

[9] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, Feb 1992.

[10] G. Zientara, P. Saiviroonporn, P. Morrison, M. Fried, S. Hushek, R. Kikinis, and F. Jolesz, "MRI monitoring of laser ablation using optical flow," *Journal of Magnetic Resonance Imaging*, vol. 8, no. 6, pp. 1306–1318, 1998.

[11] S. Roujol, M. Ries, B. Quesson, C. Moonen, and B. Denis de Senneville, "Real-time MR-thermometry and dosimetry for interventional guidance on abdominal organs," *Magn Res Med*, vol. 63, no. 4, pp. 1080–1087, 2010.

[12] B. Horn and B. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, pp. 185–203, 1981.

[13] P. Cachier, J. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis, and N. Ayache, "Multisubject non-rigid registration of brain mri using intensity and geometric features," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*. Springer, 2001, pp. 734–742.

[14] A. Azar, C. Xu, X. Pennec, and N. Ayache, "An interactive hybrid non-rigid registration framework for 3D medical images," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. IEEE, 2006, pp. 824–827.

[15] S. Loncarič and T. Macan, "Point constrained optical flow for LV motion detection," in *SPIE Proceedings of Medical Imaging*, vol. 3978, 2000, pp. 521–529.

[16] A. Becciu, H. van Assen, L. Florack, S. Kozerke, V. Roode, and B. ter Haar Romeny, "A multi-scale feature based optic flow method for 3D cardiac motion estimation," in *SSVN*, 2009.

[17] T. Brox, C. Bregler, and J. Malik, "Large displacement optical flow," *CVPR*, vol. 0, pp. 41–48, 2009.

[18] A. Sotiras, Y. Ou, B. Glocker, C. Davatzikos, and N. Paragios, "Simultaneous geometric-iconic registration," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 676–683, 2010.

[19] S. Roujol, J. Benois-Pineau, B. Denis de Senneville, B. Quesson, M. Ries, and C. Moonen, "Real time constrained motion estimation for ECG-gated cardiac MRI," in *ICIP*, 2010, pp. 757–760.

[20] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[21] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proceedings of The Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[22] A. M. Tekalp, *Digital video processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[23] I. Pratikakis, C. Barillot, P. Hellier, and E. Mémin, "Robust multiscale deformable registration of 3D ultrasound images," *International Journal of Image and Graphics*, vol. 3, no. 4, pp. 547–566, 2003.

[24] S. Roujol, M. Ries, C. Moonen, and B. Denis de Senneville, "Automatic nonrigid calibration of image registration for real time mr-guided hifu ablations of mobile organs." *IEEE transactions on medical imaging*, vol. 30, no. 10, p. 1737, 2011.

[25] *NVIDIA CUDA Compute Unified Device Architecture - Programming Guide*, 2007.

[26] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[27] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons using itks finite difference solver hierarchy," *The Insight Journal*, 2007.

[28] M. Ries, B. Denis de Senneville, S. Roujol, S. Hey, G. Maclair, M. Kohler, B. Quesson, and C. Moonen, "Three dimensional motion compensation for real-time MRI guided focused ultrasound treatment of abdomainal organs," in *ISTU*, vol. 1215, 2010, pp. 239–42.

[29] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn Res Med*, vol. 42, no. 5, pp. 952–62, Nov 1999.

[30] L. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[31] B. Denis de Senneville, C. Mougenot, and C. Moonen, "Real time adaptive methods for treatment of mobile organs by MRI controlled high intensity focused ultrasound," *Magn Res Med*, vol. 57, no. 2, pp. 319–30, 2007.

[32] M. Ries, B. Denis de Senneville, S. Roujol, and C. Moonen, "Latency compensation for real-time 3D HIFU beam-steering on moving targets," in *ISMRM*, 2010.

[33] B. Denis de Senneville, K. Noe, M. Ries, M. Pedersen, C. Moonen, and T. Sorensen, "An optimised multi-baseline approach for on-line MR-temperature monitoring on commodity graphics hardware," 2008, pp. 1513–1516.

[34] C. Rabrait, P. Ciuciu, A. Ribes, C. Poupon, P. Le Roux, G. Dehaine-Lambertz, D. Le Bihan, and F. Lethimonnier, "High temporal resolution functional mri using parallel echo volumar imaging," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 744–753, 2008.

[35] M. Black and P. Anandan, "A framework for the robust estimation of optical flow," in *Computer Vision, 1993. Proceedings., Fourth International Conference on*. IEEE, 1993, pp. 231–236.

[36] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2432–2439.

Response to the reviewers

The authors would like to thank the reviewers for their insightful and thorough remarks in helping to improve the clarity and the value of the manuscript. The manuscript has been extensively reworked according to the reviewers' comments. In particular, new synthetic experiments were performed to precisely assess the performance of the algorithms in function of their free parameters. To better assess the results, several quality criteria have been added such as the angular error and the harmonic energy. Finally, the manuscript was extensively reworked to maintain a reasonable length, given the additional results and comments added from the reviewers' comments.

Reviewer #1

The paper presents and evaluates a constrained optical flow registration algorithm. Results are convincing and tackles an often encountered problem: An application which operates under real-time constraints can only acquire two-dimensional images. Optical flow based registration of such images normally suffers "beyond recovery" from motion of objects in and out of the imaging plane, but the proposed technique seem to overcome this using "enforcement" of landmarks matching inherent in the optical flow cost function design. I can easily see their results relevant to a number of imaging applications.
The manuscript is well written and I have only a few clarifications that I would like to see incorporated in the manuscript before publication.

1.1 is there any conceptual differences between using the term "motion correction" (as you do) and tracking. You seem to resample your 'moving images' into the 'fixed image' to enable pixel index comparisons over the temporal dimension. Couldn't we just store the vector fields and track "particles" over time (tracking)?

The authors believe that both approaches are very similar. In tracking, each pixel/particle would be projected on the current image according to the current estimated motion field and an interpolation should be performed to retrieve the correct intensity at this sub-pixel precision location. The registration would just store these result as an additional step. The reviewer is right that the registration may not be necessary for such temporal analysis. However, the registered image may be used in real time for visual assessment of the algorithm in order to potentially detect unexpected behavior and eventually stop the intervention.

1.2 The sentence leading to ref. [9] p.1 c.2 is not grammatically meaningful to me

The initial sentence

"The fundamental method proposed by Horn and Schunck (referred to as H&S in the scope of this paper) introduced additional physical constraints by assuming that the motion field is smooth in the neighborhood of estimation point [9]"

has been rewritten as follows (see Page 1):

"The method proposed by Horn and Schunck (referred to as H&S in this paper) introduces additional physical constraints enforcing the smoothness of the motion field [12]."

1.3 When you introduce the concept of reduced FOV imaging at the end of page 1 and into page 2, it was (at the time of first reading it) not clear to me whether you were indeed focused on solving this specific problem in your paper or you were going through other related work. Maybe you should make clear "up front" (early in the introduction) the imaging conditions you are operating under and then review the related literature. With the current chronology the introduction seemed at bit long, but can probably be contributed to the fact that it wasn't clear exactly what the paper focus would be until the concluding paragraph in the introduction.

The authors agree that the introduction had to be clarified. The main strength of the algorithm is to provide robust motion estimation in presence of transient structures; although this can be the case in presence of out-of-plane motion, it may also represent a more general application when using reduced field of view imaging on moving structures. However, due to the over length of the paper and the requested additional results from the reviewers, the authors have decided to focus the presented paper on reduced field of view imaging applications. Therefore, the introduction has been rewritten in part and reordered to clarify the main focus of this work. In particular the paragraph related to reduced field of view imaging has been moved just after the paragraph presented the general context of the work.

The initial paragraph introducing reduced FOV imaging:

"Typically, interventional procedures are usually restricted to a part of the organ/tissue under study making reduced field of view (FOV) imaging desirable. This would allow improvement of the spatial and / or temporal resolution in order to decrease partial volume effects (undesirable for quantitative analysis) and to increase imaging framerate (required to observe rapid phenomena). Several strategies have been proposed to achieve a reduction of the field of view such as saturation slabs, which can be set around the imaging targeted area [10]. Alternatively, outer volume

suppression [11] can be used or more recently, the transmit sense technology [12] would also allow the acquisition of reduced FOV. For the particular application of cardiac function analysis and guidance of interventional procedures, Schaeffter et al. [13] also proposed a strategy for interactive reduced FOV imaging.
Although a reduced field of view may improve spatial or temporal resolution, it introduces a new challenge for the target motion estimation. Indeed, structures close to the target that would appear similar in all images using full FOV imaging (since FOV is usually centered on the target), may appear transient in the case of reduced FOV due to the respiratory motion and the limited spatial coverage."

has been modified to (see Page 1)

"Typically, interventional procedures are restricted to a part of the organ/tissue under study making reduced field of view (FOV) imaging desirable. This would allow improvement of the spatial and / or temporal resolution in order to decrease partial volume effects (undesirable for quantitative analysis) and to increase the imaging framerate (required to observe rapid phenomena). Several strategies have been proposed that use saturation slabs [4], outer volume suppression [5] or interactive reduced FOV imaging [6]. However, a reduced FOV may introduce new challenges for the target motion estimation since structures may appear transiently due to the respiratory motion and the limited spatial coverage."

A sentence has been added to the abstract of the paper as follows (see page 1):

"Reduced field-of-view imaging represents a promising way to improve spatial and / or temporal resolution. However, it introduces new challenges for target motion estimation since structures near the target may appear transiently due to the respiratory motion and the limited spatial coverage."

1.4 The overview provided in Fig. 1. and the corresponding text serves the paper well. However please include a sentence stating that a more thorough description and references for the individual components will be given later. I started wondering about many "methodological considerations" at this point of reading that you answered in the following sections.

Due to the material added to the manuscript and the limited length of the paper, the authors were thinking to substantially reduce this introduction part. However, since the reviewer found that this introduction serves well the paper, this introduction was rewritten to reduce the information to the minimum required to have a good flavor of the presented algorithm. As also suggested by the reviewer, a sentence has been added to mention that detailed descriptions of the algorithm were presented in the next sections.

The initial paragraph:

"The proposed CME algorithm is a two-step procedure (see Fig. 1). During a preparation step, constraint points are selected (step 1 in Fig. 1). For this, a mask

manually set around the target (heart or kidney in our case) in the reference image is drawn. The drawing of the mask represents the only interaction of the staff physician during the intervention, since the rest of the process is fully automatic. The edge of the mask is extracted and sampled. To refine the positioning of each sample point, a feature point detection is realized on the reference image. Then, the closest feature point for each sample point is selected. In a second step (lower block in Fig 1), the motion is estimated for each image as follows: A global translational motion estimation is performed and used to initialize a local estimation of the displacement of constraint points. Non-physiological constraint point displacements are automatically identified and corresponding constraint points are discarded. The displacements of the constraint points are then integrated into the constrained optical flow algorithm (using the global estimated motion as preconditioning) to obtain the final motion field. Due to technical limitations of fast MR acquisition sequences, extensive 3D volume imaging on mobile organs is hard to achieve. The proposed technique has thus been evaluated in 2D case in the scope of this paper."

has been changed to (See Page 2):

"The proposed CME algorithm is a two-step procedure (see Fig. 1). The first step consists of selecting the constraint points along the boundary of the organ in the reference image of the time series. In a second step the motion is estimated for each image as follows: a global translational motion estimation is performed and used to initialize a local estimation of the displacement of constraint points. Nonphysiological constraint point displacements are automatically identified and corresponding constraint points are discarded. The displacements of the constraint points are then integrated into the constrained optical flow algorithm (using the global estimated motion as preconditioning) to obtain the final motion field. A detailed description of each algorithm step is presented in the next sections."

1.5 Is edge extraction equivalent to "manual ROI specification" in the text? It is not clear to me if you do any edge detection based on the ROI or of the ROI is the edge extraction that you refer to.

The authors agree that a clarification is needed on this point. The manual ROI drawing provide a binary mask equal to 1 inside the contour of the closed manual segmentation and 0 elsewhere. The edge extraction consists in extracting the boundary of this mask. Therefore Figure 1 appeared incorrect in this sense since the figure above "mask" on the left part of the figure does not represent the real mask but the boundaries of the mask. Therefore, Figure 1 has been modified to be coherent with the employed terms where the word 'mask' has been replaced by 'ROI' and the ROI boundary has been replaced by the ROI itself. (See page 2)

1.6 you state that your iteration is based on Gauss-Seidel. Normally a GPU implementation would use Jacobi iterations due to parallelization considerations

The reviewer is right. Actually both Gauss Seidel and Jacobi have been implemented for CPU and GPU implementation, respectively. However, since we only report the GPU implementation in the paper, this has been corrected accordingly.

In the material & method part, the initial sentence:

"Finally, based on the Gauss-Seidel method, the system can be solved with the following iterative scheme:"

has been changed to (see page 3)

"Finally, based on the Jacobi method, the system can be solved with the following iterative scheme:"

1.7 for the cost function specification itself. You register all frames to a reference frame? Why do you prefer this strategy over a frame by frame concatenation of deformations? In the latter scenario every registration would probably be over more similar images and thus better matching the underlying assumption of overall image intensity conservation of infinitely small time steps.

The reviewer is partially right. This is true that, when high temporal frequency acquisition is applied the difference between two successive images is much smaller compared to the difference with the reference image. In this case, the Horn & Schunck algorithm is expected to be much less disturbed by the presence of transient structures. However, this type of strategy accumulates error over time. In our case, if a framerate (for the kidney, especially for high intensity focused ultrasound (HIFU) intervention) of 10-15 Hz is employed, this can lead to a large number of images through the entire intervention (30s-several minutes) which can generate a high accumulated error. This is the reason why the choice of a registration to a reference image has been made [1,2].

[1] de Senneville BD, Mougenot C, Moonen CT., Real-time adaptive methods for treatment of mobile organs by MRI-controlled high-intensity focused ultrasound., Magn Reson Med. 2007 Feb;57(2):319-30.

[2] Roujol S, Ries M, Quesson B, Moonen C, Denis de Senneville B. Real-time MR-thermometry and dosimetry for interventional guidance on abdominal organs. Magn Reson Med. 2010 Apr;63(4):1080-7.

The problem is slightly different in the heart since the cardiac contraction limits the framerate to the cardiac cycle (unless one can correct the cardiac contraction in real time). In this case, the time difference between two successive images is of a cardiac cycle (~1s), which mean that their difference can be very important in term of respiratory motion, and in this case a frame by frame may not help in term of intensity similarity (but will still accumulate errors).

A sentence has been added to the material and method part as follows (See page 4):

"Registration to a reference frame has been preferred to a concatenation of frame to frame estimations since the latter generally leads to the accumulation of errors, especially when a high framerate is applied such as in abdominal imaging (10-15 Hz). In cardiac applications, since the imaging framerate is generally limited to the cardiac frequency, the respiratory motion between two successive images generally depicts a large amplitude, devaluing the benefit of a frame to frame motion estimation."

1.8 Fig 3g is missing the colors indicating which graph is which (unfortunate since it is the first use)

Figure 3 has been removed and new figures have been generated.

Reviewer #2

This manuscript proposed a novel approach to robust real time image registration of mobile organs in interventional MRI. This question is of particular interest in non-invasive surgical intervention such as MR-guided thermal ablations to properly quantify the local temperature and monitor the intervention.

The proposed methods actually extends the quadratically regularized optical flow method derived by Hornst & Schunk by adding an hybrid regularization term. This term relies on the prior computation of constraint points that then enter in a distance function to penalize large displacement errors. The proposed method is supervised in the sense that the hyper-parameters involved in the regularized criterion are selected by hand by the operator. The authors propose a very convincing validation on synthetic datasets and perform a critical evaluation on a large database of in vivo experiments (kidney, heart). Also, they demonstrate that their approach outperforms the state of the art in different circumstances. Last but not least, the authors propose a GPU-based real-time implementation that makes this work valuable not only on the methodological side but also on the practical one. While some technical points (see below) could be improved, I recommend this manuscript for publication in the IEEE Trans. Of biomedical engineering after minor revisions. The points to be addressed in the following are organized in descending importance level:

2.1 Your minimization algorithm corresponds to the Gauss-Seidel coordinate descent method. This approach is known to be suboptimal (in terms of

convergence speed) in many cases even for convex criteria. Why have you resorted to such a crude minimization scheme? Given the CUDA library and the real time constraint, would it be possible to implement conjugate gradient methods or majorize minimize strategies? See for instance E. Chouzenoux et al, A majorize-minimize strategy for subspace optimization applied to image restoration, IEEE Trans. Image Processing, vol. 20, no. 6, 1517-1528, juil. 2011, or C. Labat et J. Idier, Convergence of conjugate gradient methods with a closed-form stepsize formula, J. Optim. Theory Appl., vol. 136, no. 1, 43-60, jan. 2008.

The authors agree that more optimal convergence scheme could be employed. However, the employed approach has been observed to converge between 50 to 100 iterations. Therefore, in combination with a GPU implementation, this approach required 7 ms which is perfectly suitable for our application and the imposed real time constraint. But the reviewer is right that a further investigation on the optimization scheme may be of interest to further reduce the computation time and would benefit to reduce the latency. Some works have been published to enable such optimization approach on GPU such as in

[Ali Cevahir, Akira Nukada and Satoshi Matsuoka,Fast Conjugate Gradients with Multiple GPUs Lecture Notes in Computer Science, 2009, Volume 5544/2009, 893-903,]

However, the authors did not investigate the real benefit on such approach with the proposed algorithm and would rather abstain from any speculation about the potential gain in speed using any specific method. Nonetheless, as suggested by the reviewer, a sentence has been added to the discussion part to point up the potential benefit of investigating more complex optimization approach, as follows: (See Page 8)

"Further reduction of the latency may be obtained by investigating the use of more complex optimization schemes aiming to improve the convergence speed of the algorithm."

2.2 All your dynamic datasets have been acquired using an 2D EPI sequence (single or dual shot). In the first case, you resort to parallel imaging, namely SENSE acceleration and reconstruction. Could you improve this acquisition setup by making use of the fast 3D EVI sequence, which is more robust to motion artifacts? See C. Rabrait, et al: "High temporal resolution functional MRI using parallel echo volume imaging". Journal of Magnetic Resonance Imaging, 2008; 27(4):744-53. and please comment. in your opinion, is this acquisition scheme compatible with real time motion correction?

The reviewer is right that improving the spatial coverage with a 3D acquisition is indeed of great interest since it may open the possibility to perform 3D motion

correction and may thus contribute to prevent errors arising with out-of-plane motion. Indeed, 2D heart imaging combined with slice tracking often suffers from such artifact (out of plane motion) since the navigator is positioned on the diaphragm and a relation of 0.6 between the diaphragm and the heart motion is generally assumed. However, this value is patient/volunteer dependant and may show important variations [1]. Therefore a pure 3D acquisition may help to get rid of this problem.

[1] Wang Y, Riederer SJ, Ehman RL., Respiratory motion of the heart: kinematics and the implications for the spatial resolution in coronary imaging. Magn Reson Med. 1995 May;33(5):713-9.

3D acquisitions prolong the scan time by providing much longer TR which renders the sequences much sensible to intra-scan motion and may alter the fat/blood signal cancellation. Although EVI represents a promising technique to reduce the TR and improve robustness against intra-scan motion (compared to other 3D sequences), it is also associated with new challenges that are now discussed:

- First, EVI is generally associated with a lower spatial resolution. The low resolution will increase partial volume effects which is likely to introduce a significant bias on quantitative analysis. Indeed, this would represent a high limitation for the direct application of EVI in thermometry application, especially in the heart (and in atrium) where the thickness of the wall is around 2-3 mm.

- Second, a preparation step (fat or blood suppression) is generally required prior the acquisition in the presented applications. Therefore, the potential regrowth of the fat signal or the efficiency of the blood signal suppression should be investigated in such conditions since they may also hamper the quality/value of the data.

- Finally, for the specific case of PRF based MR-thermometry, the k-space center should be sampled at TE=T2* for an optimal sensitivity to the temperature induced phase shift. Therefore, the k-space sampling employed by the EVI sequence (around 200 ms in the paper cited by the reviewer) should be done in a way to first acquire the k-space center (spiral acquisition?).

Therefore, the authors believe that EVI may represent an interesting technique toward a 3D extension of the acquisition. However, we also believe that a direct application of such method would not be straightforward since it would provide new questions/challenges to address (how to decrease the spatial resolution, what is the impact of a longer readout time respect to fat or blood suppression? and what would be the optimal k-space sampling?). However, as suggested by the reviewer a paragraph has been added to the discussion part to draw the potential interest/limitation of 3D acquisition and EVI as followed: (See page 8)

"Due to technical limitations of fast MR acquisition sequences, extensive 3D volume imaging on mobile organs is hard to achieve. The proposed technique has thus been evaluated in the 2D case. An extension of the method towards 3D motion correction may also be considered in future works and may improve the motion correction in the third dimension by reducing out-of-plane motion artifacts. However, although the proposed algorithm can be easily extended to 3D, the main challenge would likely remain in the design of a reliable 3D MR-sequence. In such acquisitions, the scan time increase would render the sequence more sensitive to intra-scan motion and may alter the fat and the blood signal cancellation. Echo volumar imaging [34] may help to decrease the scan time of such 3D acquisition; however, its associated low resolution and its robustness against the latter artifacts should be carefully investigated."

2.3 Your feature point detection is based on the Harris-Stephens detector ie Eqs (3)-(4) in which a sensitivity parameter k enters. Please comment on how did you select a fair value for k? In the same vein, in Subsection II.C, you set the different hyper-parameters to given values that remain the same for all methods. How did you choose them? Since I really appreciate your sensitivity analysis with respect to N and R in Fig. 8, I would be very interested in measuring the effect of varying alpha and lambda. Could you perform this additional study?

Choice of the sensitivity parameter k:

The value of the sensitivity parameter k employed in equation 3 and 4 has been taken from literature value range (0.04 to 0.15). A value of 0.04 was empirically found suitable in our case. Although an individual calibration of this value may provide better results in terms of feature point characterization, the employed value always provided coherent feature point (right on the border of the organ). Therefore, the potential gain achievable by a fine tuning of this parameter appeared more of theoretical than of practical interest. Therefore, a precise tuning of this parameter did not appear as the key point of the proposed algorithm.

A sentence has been added to the material and method section to better describe the choice of the k value as follows (See page 3)

"The value of k is generally chosen in the [0.04-0.15] interval. A value of 0.04 was determined empirically suitable for our application."

Influences of the motion estimation parameters

Since the reviewer #3 also pointed out the interest of a more complete evaluation of the algorithm sensitivities with respect to its parameters, we decided to emphasize this aspect in the revised version of the manuscript. Although such comprehensive study can be performed (as we did in the first version of the paper) on the in-vivo

datasets, the results are limited by the precision of the computation of the DICE coefficient (which is based on a manual drawing of the registered target). Therefore, it becomes more complicated to depict/distinguish small variations. Consequently, the authors decided to perform a complete new comprehensive study on a new dataset experiment where the real motion was know and thus where the true error could be computed. The sensitivity of the proposed algorithm was studied against the following parameters (α^2, λ^2, R^2, N). Endpoint error (EE, GSE in the initial version of the paper), harmonic energy and angular error are reported in Figure 2.

The initial description of the synthetic experiment in the material & method section:

"Synthetic dataset experiment: A sequence of T(= 30) images was created. To simulate respiratory motion typically encountered on mobile organs, a "ground truth" periodic translational displacement field Dgt(t) (maximum amplitude=7.5 pixels, step size=1.5 pixels) was synthesized. The signal-to noise ratio (SNR) of 20 was chosen to simulate a realistic acquisition (typically between 5 and 20). To simulate the presence of a structure appearing transient, a rectangular structure was added in half of the images (corresponding to the images with the three largest displacement amplitudes of each cycle)."

has been changed to (See page 4):

"Synthetic dataset experiment:

Data creation: A sequence of T(= 30) images was created. To simulate respiratory motion typically encountered on mobile organs, a periodic (period=6 frames) geometric transformation composed by a 2D translation (T_x,T_y) and scaling (S_x,S_y) was synthesized (T_x = {0, 0.5, 1, 1.5, 1, 0.5}, T_y = {0, 2.5, 5, 7.5, 5, 2.5} pixels, S_x = S_y = {1, 1.03, 1.06, 1.09, 1.06, 1.03}). A signal-to-noise ratio (SNR_{dB}) of 1.3 was chosen to simulate a realistic acquisition (typically between 0.7 and 1.3). A structure appearing transiently in the lower part of the image was added in half of the images to simulate the effect potentially encountered with reduced FOV imaging. Quality assessment of the motion estimation: Since the real motion (D_{gt} = (u_{GT}, v_{GT})) and the estimated motion (D = (u, v)) are available for each pixel in such synthetic dataset experiment, the measures commonly reported in the optical flow community such as the endpoint error (EE) and the angular error (AE) of the flow [9], [26] were computed, with:

$$EE = \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2} \qquad (11)$$

$$AE = \cos^{-1}\left(\frac{1 + u \cdot u_{GT} + v \cdot v_{GT}}{\sqrt{1 + u^2 + v^2}\sqrt{1 + u_{GT}^2 + v_{GT}^2}}\right) \qquad (12)$$

In addition, to provide additional information about the smoothness of the estimated motion field, the harmonic energy of the estimated flow [27] was computed."

In the Results section, the initial figure 2 has been replaced by two new figures (Figure 2,3). The initial synthetic result description:

"The comparison between our method and the H&S algorithm was first realized on a synthetic dataset where the reference image and an image corresponding to a different position of the motion cycle are displayed respectively in Fig. 2a and b. The appearing rectangle, located at the bottom of Fig. 2b, is expected to hamper the motion estimation at the bottom of the object (in the red ROI displayed in Fig. 2a). Therefore, the gold standard error was computed for each dynamic, over this ROI and the results are plotted in Fig. 2c. The H&S algorithm showed very poor performance on this area where the appearing rectangular structure biased the accuracy of the algorithm (three repetitive high values). On the other hand, the proposed CME approach remained stable over the time and provided a more accurate motion field."

has been changed to (See page 5)

"An example of synthetic images is shown in Fig. 2 where the reference image (Fig. 2a, containing the transient structure) and an image corresponding to the maximum synthesized motion (Fig. 2b, 4th image of the cycle) are displayed. The contours (interior and exterior) of the target obtained from the reference image are reported in dashed red curves. The H&S approach was employed to register the image in Fig. 2b using different α^2 values. When using a small value (α^2 = 0:5), the registration is strongly influenced by the underlying structure leading to a severe registration artifact (yellow arrows in Fig. 2d). By using an optimally calibrated value (α^2 = 2:5, see Fig. 3a), the registration artifact was reduced but still present (yellow arrows in 2e). A higher value of α^2(= 10) is reported in Fig. 2f where the registration artifact introduced by the transient structure appeared substantially reduced due to a higher weight on the smoothness of the motion field. However, in this case a different registration artifact was observed due to the inability of the method to handle complex motion (here the scaling effect) as shown by the yellow arrows. The CME approach provided a reliable registration where the registered target perfectly matched the reference target contour. These results were confirmed in the plot of the averaged EE values over time (computed inside the target, between the two red dashed curves) in Fig. 2g. While the H&S approach periodically failed to recover the correct motion, the CME offered more robust performances over time.

In order to better characterize the potential gain and limitations of the compared methods, their performance in terms of averaged error endpoint, harmonic energy and averaged angular error were precisely investigated as a function of the employed parameters (Fig. 3). As previously observed in Fig. 2d, low α^2 values (Fig. 3a, 3e and 3i) provided poor performance since the methods became very sensitive to the presence of the intrusive structure. On the contrary, high values limited the ability of the algorithm to estimate complex motion and also deteriorated the motion estimates as confirmed by the convergence of the harmonic energy toward a very small value. The averaged EE values obtained with an optimal α^2 calibration were 0.42 for the

H&S approach (α^2 = 2:5) and 0.21 for the CME method (α^2=6.5) showing a reduction of the averaged EE by a factor of 2. The λ^2 value influence was then investigated ((Fig. 3b, 3f and 3j). As expected, small λ^2 values tend to the H&S performance and high values tend to the extrapolation of the constraint point motions (and their associated uncertainty) leading in both cases to a deterioration of the registration. A good calibration of λ^2(=0.1, blue curve) provided a significant improvement of the motion estimation (Fig. 3b, 3j). The influence of the constraint point number (N) was then evaluated (Fig. 3c, 3g and 3k). Although an optimal value was reached around 20 points (Fig. 3c, 3k, blue curve), the sensitivity of the CME to this parameter was limited. Finally the influence of the bandwidth (R^2) provided an optimal calibration for a value of 5 (Fig. 3d, 3h, 3i, blue curve)."

Finally, while these results were not shown in the initial submission of the paper, a discussion had been already integrated to the manuscript. Therefore, this part was conserved in the revised version of the paper.

Choice of the motion estimation parameters

The employed parameters for the in-vivo study were initially calibrated based on the registration performance obtained on similar data acquired without reduced field of view imaging. The authors agree that some clarification is needed on this point and should be added to the manuscript.

The parameters calibration of a motion estimation algorithm is difficult in-vivo since the real motion is unknown. Especially, by considering the initial synthetic case, as correctly pointed out by the reviewer #3, a higher α^2 value will increase the smoothness of the motion and will reduce the registration artifact introduced by the intrusive structures. However, such configuration will also decrease the ability of the algorithm to estimate complex motion. We recently showed in

S. Roujol, M. Ries, C. Moonen, and S. Denis, "Automatic nonrigid calibration of image registration for real time mr-guided hifu ablations of mobile organs." IEEE transactions on medical imaging, vol. 30, no. 10, p. 1737, 2011.

that an optimal range of α^2 values for the correction of the respiratory motion on abdominal imaging was around [0.1-0.5] and a plateau was generally reached within this range of values. We also pointed out in the discussion that using the lower range of such plateau may help to improve the robustness of the algorithm against motion not encountered during the calibration step (with a potential higher amplitude, due to a deeper breath for example).

Therefore, the choice of an optimal α^2 value is the presented reduced FOV imaging case appears more tedious. The optimal α^2 value (0.1-05) obtained from full FOV imaging will provide an optimal registration in all our target except in the area closed

to the intrusive structure. On the other hand, a higher value of α^2 will decrease this artifact but will deteriorate the registration in areas subject to complex deformations (which for the case of the heart may represent the whole myocardium due to its complex deformation). Therefore, the optimal calibration depends on the criterion to be optimized. Therefore, we believe that in term of clinical perspective it is more valuable to provide an algorithm working correctly in the majority of the organ (using a smaller α^2 value calibrated on full FOV images, but potentially leading to strong local registration artifact) than a higher value which might decrease the overall DICE value while affecting/degrading the majority of the registration (due to its inability to cope with complex motion).

A paragraph has been added to the material and method part to justify this point as follows (See page 4)

"In-vivo calibration of the employed algorithms is complex since it depends on the criterion to be optimized. The calibration of the H&S method (α^2 value) has to deal with contradictory effects. A small regularization of the motion field is required to enable the estimation of complex motion and to have a globally reliable estimated motion. On the other hand, a high constraint on the motion field smoothness would reduce the registration artifact induced by intrusive structures but at the same time will limit the ability of the algorithm to estimate complex motion. This may thus deteriorate the registration in the entire organ. Therefore, we calibrated the employed algorithms in a way to maintain optimal performance in the general case of full FOV imaging (without the presence of intrusive structures). As recently shown [24], a reliable in vivo calibration of the H&S algorithm was obtained for a range of α^2 values between 0.1-0.5. A plateau was generally observed for these ranges of values and its lower bound was suggested as a good way to cope with variations of the breathing pattern (such as an amplitude variation or drift of the respiration pattern). Therefore, a α^2 value of 0.1 was employed for the H&S algorithm. We empirically found a near-optimal solution for the CME calibration by employing the following parameters: α^2=0.1, λ^2=0.1, N=20 and R^2=5. Note that a similar optimal configuration was obtained for the synthetic dataset experiment (see Result section) except for the α^2 values that were higher due the lowest complexity of the synthesized motion."

2.4 Please comment also why have you retained a Gaussian rho function in Eq. 6? What about the presence of outliers and the putative relevance of more robust metrics (Huber function instead of d^2)?

The authors agree that the employed metric used in equation 6 is not robust against outliers. However, this Gaussian rho function is applied on landmark points that have been already tested for the presence of outlier (3 sigma rules applied on landmark) and eventually rejected if detected as outlier. However, the authors believe that such robust metric may be of interest for equation 5 (especially for the two first terms) instead of using L2 norm, as initially proposed in the robust formulation of optical flow

by Black and Anandan [1]. This may improve to account for motion discontinuities and for both brightness and smoothness violation. Therefore, a sentence and the corresponding reference have been added to the discussion part of the manuscript about the potential gain provided by a more robust metric as follows (See page 8)

"A robust formulation of equation 5 using robust estimators such as the Humber function or Lagrangian [35], especially in the first two terms of equation 5 should be investigated in order to account for both brightness variation and smoothness violation (motion discontinuities)."

[1] Black, M. J. and Anandan, P., **A framework for the robust estimation of optical flow,** *Fourth International Conf. on Computer Vision*, ICCV-93, Berlin, Germany, May, 1993, pp. 231-236

2.5 Please report SNR measure in dB in Subsubsection II.D.1, page 3, col1, last paragraph.

SNR measures are now reported in dB as suggested by the reviewer.

2.6 Figures are in general too small. Please use the figure* environment instead of figure in Figs. 3,5,6 and enlarge your images. Also, remove the Matlab x- and y-labels and crop the figure to magnify them. Put the axis legend in the core of the text (Figs 3-8).

The authors agree that most of the figures were too small and made an effort to enlarge the new figures. However, since the new figures are rather dense, we decided to keep the x and y labels to improve their readability.

Reviewer #3

The paper presents a method for pairwise registration that enhances the Horn & Schunk method by considering an additional regularization term that takes into account pre-estimated point correspondences.

The paper is well organized though the writing could be improved greatly. The writing could profit greatly from the reading of a native English speaking editor as there is an important number of grammatical and syntax errors. To point some of them, the following notation is going to be used: P2L10l = page 2, line 10, left column.

3.1 Some general comments:

All the following corrections have been implemented in the text. When required the answer/modification has been reported in the following.

i) there is an important problem with the use of definitive and indefinite articles.
ii)Equations should be treated as a single noun and they should be followed by a comma, especially when a description of its individual members follow (which is the case for most if not all equations in the paper). If an equation is at the end of a sentence, a period should follow the equation (equation 1).
iii)The use of "dynamic(s)" is neither proper nor clear.

Dynamic(s) has been corrected in frame(s).

iv)After some equations (Eq. 1, 9, 10) a new paragraph starts whereas it shouldn't.

P1L23l: "allows integration of displacement of physiological landmarks, which are obtained in a preparation step" -> In general this sentence is problematic. Articles are missing: allows the integration of... is more correct. "of displacement of physiological landmarks" does not seem right. preparation is a noun, preparatory should be used instead.

P2L19L: "the image to register" -> image to be registered.
P2L23l: "increase α^{2}" -> increase the α^{2}
P2L24l: "in classical" -> in the classical
P2L50l: "to its initially" -> its initially
P2L13r: "paper, where preliminary results": The sentence needs rephrasing. The use of "where" is wrong here. Maybe split it in two sentences.
P2L16r: "FOV imaging and with out-of-plane" -> FOV imaging with out-of-plane
P2L20r: "estimation as a complementary regularization term" -> estimation by introducing an additional regularization
P2L22r: "constraint point influence" : I find the use of nouns as adjectives tiring to follow as it is difficult to see which is the "real" noun. In my opinion "local control of the influence of the constraint points" is preferable.
P2L44r:"The edge of the mask" I believe the use of contour instead of edge is better.
P2L49r: "as follows: A" -> as follows: a
P2L60r: "evaluated in 2D case" -> evaluated in the 2D case
P3L20l: "tracked over the time" -> tracked over time
P3L28l: "then regularly spaced sampled" -> sampled regularly in space
P3L36l: "Typically a small neighborhood of regularly spaced sampled point is chosen to search for a closest feature point (e.g. 3 $\times$ 3 pixels

neighborhood)". In general, this phrase does not help me understand how this step is performed. Could you elaborate more? Moreover, "for a closest" -> for the closest and "pixels neighborhood" -> pixel neighborhood.

This paragraph describing the landmark selection has been rewritten for clarification as requested by the reviewer in 3.3. The current sentence has been changed to *(See page 3)*

"Therefore, the feature point with the higher response in a small neighborhood (e.g. 3x3 pixels) of the contour point is selected as a landmark point."

P4L7r: "near the global optimum": I find the use of global optimum somehow misleading as no algorithm is able to attain the global optimum solution. The assumption most algorithms do is that of small displacements. Thus, the "initialized after compensating for large displacements" seems more right.

The reviewer is right. The sentence has been changed to *(See page 3)*

"An optical flow based algorithm is more efficient when it is initialized after compensating for large displacements"

P3L13r: "is restricted to a ROI defined in" -> is restricted to the ROI defined
P3L28r: "its displacement coordinated" -> displacement components
P3L55r: "the variation calculus" -> calculus of variations
P4L4l: "u in the neighborhood [9]" -> whose neighborhood?how is this neighborhood defined?

\overline{u} is the mean value of u in the neighborhood (3x3 pixels) of the estimated point. This information has been added to the manuscript as follows *(see page 3)*

"with u the mean value of u in the neighborhood (3x3 pixels) of the estimated point [12]."

P4L33l: "a multi-" -> the multi-
P4L34l: "which iterates the motion" -> which refines(?) the motion.
P4L46l: "was based on Compute Unified" -> was based on the Compute Unified
P4L58l: "The signal-to-noise" -> A signal-to-noise
P4L2r: "a structure appearing transient" -> transient structure or appearing transiently.
P4L18r: "positioned in head first in supine " -> positioned head first in supine
P4L42r: experiment except the"-> experiment except that the
P4L53r: " In such conditions" -> no new paragraph

P4L60r: "difference of motion compensated images" -> difference between what and what?

The authors were talking about the difference between the motion compensated images and a reference image. The sentence has been removed in the revised manuscript.

P5L3r: "dynamic, over" dynamic-> frame? time instance? dynamic is either adjective or a noun meaning either interactive system or a force. This goes for all cases where "dynamic(s)" is used (i.e. P5L56r, P5L58r, P6L60l, P8L59l)

Dynamics has been changed to frame in the manuscript.

P5L53r: "were obtained from the" -> were obtained for the
P6L49l: "performance in the heart of six" ->it is not the heart of the volunteers. Call it heart data set but "in the heart", in my opinion, means something really different from what you would like to say.
P6L13r: "5c,5d,5e" -> 5c, 5d, 5e.
P6L17r: "structure appearing transient" -> transient structure or appearing transiently.
P6L52r: "obtained in the heart" -> see previous comment
P6L57r: "Influences of the CME" -> Influence of the CME
P7L47l: "with closed parameter" -> with close parameter
P7L15r: the enumeration of the sub figures is wrong. There are two "d" and "e" should be "f"
P7L56r: "method on synthetic" -> method on a synthetic
P8L42l: "structure appearing transient" -> transient structure or appearing transiently.
P8L9r: "prediction algorithm performance greatly increases" -> the performance of the prediction algorithm increases greatly
P8L51r: "Outliers rejection"-> Outlier rejection
P9L5l: "optical flow formulation" -> optical flow formulations
P9L13l: "feasibility with the presence" -> feasibility in the presence
P9L4l: "This point is particularly important since many optical flow formulation using many terms usually require much more free parameters, rendering the outcome of the algorithm dependent on the user calibration"->This is not true. I can understand an argument of the type "using landmark information, our method is more robust to the choice of the parameters" but most optical flow methods that use only the intensity information need only to choose a weighting factor to balance the data and regularization terms. Most optical flow methods have less parameters than the proposed one.

The reviewer is right and the sentence has been removed.

P9L20l: "renders optical flow method" -> renders an optical flow ... or renders optical flow methods
P9L23l: "point displacement" -> point displacements
P9L25l: "of another motion" -> of other motion
P9L25l: "navigator echoes or ultrasonic echoes" -> navigator or ultrasonic echoes
P9L31l: "extensions of the method has to" -> extensions of the method have to

Thank you very much for all of these corrections. The manuscript has been corrected accordingly. In addition the manuscript has been read and corrected by a native English speaking person as requested by the reviewer.

3.2 Regarding the introduction, as the most important novelty of the proposed framework lies in the introduction of the regularization term that introduces landmark information, I would expect a more detailed presentation of related techniques that exploit both intensity and landmark information. Some of the most recent methods towards this direction that should be cited are: Aristeidis Sotiras, Yangming Ou, Ben Glocker, Christos Davatzikos and Nikos Paragios. "Simultaneous Geometric - Iconic Registration." MICCAI 2010 A. Azar, C. Xu, Xavier Pennec, and Nicholas Ayache, "An Interactive Hybrid Non-Rigid Registration Framework for 3D Medical Images", ISBI 2006 P. Cachier, J. F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis, and N. Ayache, "Multisubject non-rigid registration of brain MRI using intensity and geometric features", MICCAI 2001

The introduction part of the manuscript has been rewritten in part for clarification as also requested by the reviewer #2. The proposed references have been added. The initial paragraph in the introduction part

"To overcome the limitation imposed by the assumption of conservation of either dense feature descriptor values or intensities, Loncaric et al. [16] proposed to constrain the iterative scheme of the H&S algorithm by introducing constraint points. Such points were selected on the pre-segmented border of a moving organ and their displacements were estimated by a local block-matching with a pure translational model. Hence a sparse set of motion constraints was introduced to constrain the optical flow estimator. The algorithm [16] was designed to obtain a final motion field that exactly fits, for each constraint point, to its initially estimated displacement. This algorithm requires an a priori exact knowledge of each constraint point displacement or at least a very good estimation of the latter. However, with a deformable complex motion and inherent noise in MRI sequence, this information is generally only available as an estimate. An attempt to integrate feature points into the optical flow formulation was proposed by Becciu et al. [17] for cardiac contraction analysis using

tagged MRI. In their work, feature points were extracted from the tags. The good contrast level on tags allows for sufficiently precise detection of suitable constraint points. However, MR tagged images are generally unsuitable for interventional MRI, since images are tagged by regular lines where the signal has been removed. Recently, a variational approach, integrating segmented region motion, was proposed for large displacement estimation [18]. This method uses a linearized OFE deviation together with regularization terms which include the correspondence of region displacements in the image plane. Despite the interest of such an approach in general purpose video sequences, its application to MRI sequences is not straightforward due to the inherent difficulties of segmentation of frames into spatially coherent regions."

has been changed to (See page 2):

"Therefore, the first step of this study was to hypothesize that a combination of intensity and landmark registration could be used to improve the robustness of the registration against transient structures. Such combinations have been proposed in the past such as in [13] where a large number of landmarks were employed for brain registration. This method was further extended to deal with a smaller number of points (manually defined) for interactive registration of medical images [14]. In these approaches, both intensity and landmark metrics were minimized sequentially inside an iterative minimization process. Several unified minimization frameworks were also proposed. In [15] the algorithm was designed to obtain a final motion field fitting each pre-estimated landmark motion. In the presented application, due to the low SNR, a non-negligible uncertainty of the landmark correspondence may be expected and would thus strongly affect such registration approach. Another solution proposed by Becciu et al. [16], attempts to use tags obtained from an MR-tagged sequence in their registration method for cardiac contraction assessment. Unfortunately, this MR-sequence is generally unsuitable for interventional MRI, since images are tagged by regular lines where the signal has been removed. Recently, a variational approach, integrating segmented region motion, was proposed for large displacement estimation [17]. This method uses a linearized OFE deviation together with regularization terms which include the correspondence of region displacements in the image plane. Despite the interest of such an approach in general purpose video sequences, its application to MRI sequences is not straightforward due to the inherent difficulties of segmentation of frames into spatially coherent regions. Recently, a grid-based deformation model was proposed [18], but was also shown to be sensitive/limited by the landmark extraction process according to the authors."

[13] P. Cachier, J. F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis, and N. Ayache, "Multisubject non-rigid registration of brain MRI using intensity and geometric features", MICCAI 2001

[14] A. Azar, C. Xu, Xavier Pennec, and Nicholas Ayache, “An Interactive Hybrid Non-Rigid Registration Framework for 3D Medical Images”, ISBI 2006

[18] Aristeidis Sotiras, Yangming Ou, Ben Glocker, Christos Davatzikos and Nikos Paragios. "Simultaneous Geometric - Iconic Registration." MICCAI 2010

3.3 Regarding the methods, i find the description of the 1st step of the algorithm (constraint point selection) not clear enough. The physician traces a ROI which is then refined by moving its contour points towards the closest feature points. In what sense is the closest used here? Closest when considering their spatial distance or based on intensity similarity criterion. Moreover, I found the description of the search for the closest feature point not easy to understand. Why do you choose a neighborhood of contour points to search for the closest feature point? Don't you try to find a feature point for every contour one? Shouldn't you consider multiple feature points for each contour one? What happens if two feature points are equally close to the contour one? How do you tackle the case that that the same feature point is assigned to more than one contour points?

The authors agree that a clarification is needed. First, the term “closest” was improperly used. Indeed, the feature point having the higher response in the neighborhood of the subsampled contour point was selected (we only assign one feature point to each subsampled contour point, and not to each point from the initial continuous contour). In the case that no feature point is found, the initial contour point is selected. If multiple feature points are obtained, the one with the higher response is chosen (which is not necessary the closest one from the initial contour point).

The search of a closest feature point was restrained to the neighborhood of the contour point to prevent the positioning of a feature point on another organ/tissue, that is likely to have a different motion from the heart. In such a case, the landmark would give incoherent displacement values and may be discarded by our outlier rejector. Therefore such points would be useless to guide the registration process.

Lastly, if a feature point is assigned to two contours points, one could have chosen to use such point only one time or twice. In the current implementation, we decided to use it twice, however, this case is not relevant in in-vivo cases since the distance between two contours points was always superior to the search neighbor window for feature points (given the employed number of points N=20).

A paragraph has been added to clarify the description of the landmark selection in step 1 as follows (See page 3)

“Therefore, the feature point with the higher response in a small neighborhood (e.g. 3x3 pixels) of the contour point is selected as a landmark point. The small neighborhood search was designed to prevent positioning a landmark on a different

structure/organ that may depict a different motion from the targeted organ. If no feature point is present, the initial ROI contour point is then selected."

3.4 Regarding the constrained registration part it is rather interesting and novel. There is two things though that seem counter-intuitive. First, the influence of the number of constraint points. I agree that with few points, the accuracy of the motion estimation should be low. Nonetheless, I would expect a monotonic increase of the accuracy with an increasing number of constraint points (till probably a plateau). Could you interpret the decrease of the performance when passing from 20 to 40 points? One last thing regarding the parallel implementation of the algorithm. In the paper, it is stated that a reduction factor of 10 was achieved for the computation time of the iterative numerical scheme of the optical flow algorithm. Why the gain factor is so low? A more significant gain should be expected given the difference between the computational power of the CPU when compared to the GPU (1 or 2 cores vs 240).

Influence of the number of landmarks

We observed that a large number of landmark point do not necessary improve the registration. Indeed a large number of landmark lead to an increase of the overall weight of the landmark influence on the final estimated motion. Therefore, if the landmark motion is perfectly correct, the authors agree that a monotonic improvement of the registration performance should be observed with the increase of the number of landmark. However, if a bias (which is expected to be locally correlated for close feature points) is present on the landmark displacement, the conclusion may be different. In this case, if only a small number of landmarks is employed, the CME can still correct for this approximation by using the two first terms of equation 5. If a large number of landmarks is used, then the landmark displacement uncertainties have a broader impact on the registration process, resulting in a decrease of the registration performance.

Based on the remark 3.5 of the reviewer, a complete new synthetic dataset experiment was performed confirming this explanation. As can be observed in the Figure 3 a performance decrease towards a plateau was observed when using a high number of landmark points (see Figure 3c, 3g and 3k).

In summarize a too large number of landmark decreases the ability of the algorithm to correct for uncertainties on the estimated landmark motion. This point has been added to the discussion part as follows (See page 8)

"On the other hand, a too large number of constraint points will slow down the computational process and may degrade the registration performance by leading to a quasi interpolation of the constraint displacements and their associated uncertainties"

GPU acceleration and performance

Considering the acceleration factor, this value depends on both employed CPU and GPU. First the employed CPU was a very powerful unit. Then, by looking at the GPU implementation, previously published works [1,2] reported a computation time for the Horn & Schunck algorithm around 5-20 ms for the same image size (128x128) depending on the employed GPU card. Using the current GPU card employed for this study, the Horn & Schunck algorithm was performed in 5 ms. Therefore a total computation time of 7 ms for the proposed optical flow algorithm appeared in line with the published works.

Also the ratio of the core number may also be limited for such comparison since the frequency of each processor is not comparable (602Mz on GPU vs 3.1 Gz for CPU).

Finally, the relatively small amount of data to be processed (image of 128x128 pixels) may also contribute to such performance. As shown in [1], an acceleration factor of 7.6 was achieved on image resolution 128x128, while an acceleration factor of 26 was obtained with images of resolution 512x512.

[1] Baudouin Denis de Senneville, Karsten O. Noe, Mario Ries, Michael Pedersen, Chrit T. W. Moonen, Thomas Sangild Sørensen: An optimised multi-baseline approach for on-line MR-temperature monitoring on commodity graphics hardware. 1513-1516

[2] Roujol S, Ries M, Quesson B, Moonen C, Denis de Senneville B. Real-time MR-thermometry and dosimetry for interventional guidance on abdominal organs. Magn Reson Med. 2010 Apr;63(4):1080-7.

This has been added to the discussion part. The initial sentence:

“Here, the demonstrated CPU/GPU implementation allows the acceleration of the required processing time by a factor of four and thus ensuring real time conditions with low latency.”

has been changed as follows (See page 8)

“Here, the GPU implementation offered an acceleration factor of 10 for the computation of the constrained optical flow which is in accordance with the published work [11], [33]. A significantly higher acceleration factor would be expected using higher resolution as shown in [33] since it would benefit from a larger amount of data to process (higher occupations of each processor) and a lower relative overhead. Overall, the demonstrated CPU/GPU implementation allows the acceleration of the required processing time by a factor of four and thus ensures the real time conditions with a short latency.”

3.5 I believe that the experimental validation section can be improved greatly. My first concern is regarding the choice of SIFT-flow as a method to compare to. I was troubled by this choice as the authors of [14] argue in the recent PAMI version of their paper (that i believe should be cited instead of the conference one) that optical flow cannot be replaced by SIFT flow. The main disadvantage of SIFT flow is that it cannot provide sub-pixel accuracy. Thus, a comparison with methods that can have such a performance is not meaningful especially when the code for other state-of-the-art optical flow methods (i.e. Secrets of optical flow estimation and their principles. CVPR 2010) is available on-line. Furthermore, it is not clear how the SIFT-flow was used. More details with respect to the use of the code should be given. How was the SIFT descriptor created? The parameters that were used should be detailed.

The authors agree that a direct comparison of SIFT flow with optical flow is biased by the inability of the SIFT flow to provide a sub pixel accuracy. However, the authors decided to include the SIFT flow approach in this study to compare our method with non-rigid registration approaches using either a purely landmark based method (SIFT flow) or a purely intensity based method (Horn Schunck). The objective was to show that both purely landmark based and intensity based approaches were strongly affected by the presence of transient object. In such a case, as with the Horn & Schunck method, the SIFT flow is largely disturbed by the presence of an intrusive structure, and shows an error with a magnitude much larger than the pixel accuracy. Therefore, one can conclude that the main error obtained with the SIFT flow method in this case can be attributed to its inability to cope with such intrusive structure and not to its inability to provide sub-pixel accuracy. To be complete, the initial calibration of the SIFT flow has been initially performed following the strategy described in the answer to 3.6, where an extensive discussion about the influence of the motion estimation parameters and their calibration is given for the H&S and the CME methods.

Overall the authors agree that the SIFT flow comparison is limited and may not contribute to the clarity of the paper by providing similar conclusions as obtained with the H&S approach. Also, since additional results were requested by the reviewer in 3.6 and 3.7 (with an important quantity of new experiments/results were added to the revised manuscript), and to limit the length of paper (which is already rather long for the IEEE TITB format), the authors decided to remove the SIFT flow comparison from the paper.

<u>“Secrets of optical flow estimation and their principles”:</u>

The paper “Secrets of optical flow estimation and their principles” provides very interesting and impressive results by showing that the combination of H&S method

with modern optimization and implementation technique can provide very good results. In detail, they show that the use of a median filter on the flow after each iteration step improves significantly the registration since their algorithm ranked 1st on the Middleburry evaluation. Therefore, the median filter could be easily integrated to our algorithm and represent a promising way for further improvements of the CME performance. Since this integration can be performed in both the H&S and the CME methods, the benefits of such approach for each method represents a nice study by itself, which is, in our opinion, beyond the scope of this paper. Therefore, the authors would intent to pursue this promising idea in future work and a sentence has been added to the discussion part to comment such future extension as follows (see page 8):

"Also, the integration of a median filter at each iteration step of the iterative minimization scheme has to be carefully investigated since significant improvements have been achieved in this direction [36]. Nevertheless, the real time compatibility of such approach has to be evaluated."

3.6 Regarding the comparison that is performed in the paper I have one major comment. It is difficult to compare different methods as it is not evident if the difference in their performance is due to a better tuning of their parameters or is because of their different qualities. Thus, either each method should be tuned separately and the best results should be reported or the results should be given for various regularization parameters. Ideally, graphs of the DICE coefficient (or GSE) should be given plotted against the harmonic energy of the deformation field. That way, we can understand what is the cause of the more accurate performance. Now, for example one may argue that in the synthetic case (section III. A) the H&S method could perform better by regularizing more. At least, the values of the harmonic energy should be given for each case in order to let the reader have a better understanding of the added value of the adoption of point constraints.

The authors fully agree with this point. The tuning of a registration algorithm clearly affects the algorithm output and the resulting performance. Also the authors would like to apologize for the lack of information about the algorithm calibration/tuning. To better understand the interest of adopting the CME registration and to answer the reviewer comment two major modifications have been brought to the manuscript:

- A complete new synthetic dataset experiment has been performed in order to precisely analyze the algorithm performance in function of their (optimal and non optimal) calibrations. Endflow error, angular error (as requested in 3.7) and harmonic energy have been reported for a wide range of parameters.

- The optimal in-vivo calibration of the algorithms has been clearly detailed and the harmonic energy plot added to the results.

We now describe in detail the modification / explanation / additional experiments that have been carried out in both synthetic and in-vivo experiments

3.6.1 Synthetic Dataset experiment:

We decided to perform a complete new synthetic dataset experiment to precisely evaluate and show the influence of the CME parameters (α^2 ,λ^2, *N and* R^2) and compared the result with the H&S method (with also a complete study of the α^2 value). The interest of doing such evaluation on synthetic dataset is twofold:

- The real motion is known. Then absolute criteria such as the endpoint error (EE) of the flow (that was referred to as GSE in the previous version of the manuscript) and angular error can be computed.
- Since the assessment of the registration performance can be automated (thus do not require manual drawing as for the DICE coefficient), a large number of tests can be performed allowing for a precise sampling of each parameter space.

The endpoint error (EE) and angular error (AE) of the flow have been computed over the synthetic target (between the red dashed line in Figure 2a) area. The corresponding harmonic energy (HE) computed over the whole image have been also calculated for all tested cases. The results are reported in figure 2 and 3.

The CME clearly overcomes the H&S algorithm as shown by the optimal EE, AE and HE values as shown in figure 3a, 3e and 3i. As pointed out by the reviewer, the increase of the α^2 value reduces the registration artifact induced by the intrusive structure until reaching an optimal configuration. Then, higher α^2 *values start degrading the performance by enforcing a high smoothness constraint preventing the estimation of the synthesized complex motion. Note that the optimal* α^2 *is expected to be much lower in in-vivo dataset since a more complex motion should be encountered in-vivo. In the* λ^2 *analysis, it is interesting to note that while a small* λ^2 *values tends to provide an H&S like behavior, a high value tends to extrapolate the constraint motion over the target. One can observe in the latter case that the "constraint motion extrapolation registration" provide an EE which can be substantially improved by using an optimal* λ^2 *value of 0.1. This is explained by the uncertainty/bias of the constraint motion that can be corrected using smaller* λ^2 *values, but which is extrapolated using higher* λ^2 *values. Also, by increasing the number of points N (to 20), one can improve the registration performance. However, this is not longer the case for a higher values of N where the weight of the landmark become higher in the registration process tending again to extrapolate the constraint motion (and its associated uncertainty) over the target. Finally, the influence of the bandwidth (*R^2*) provided an optimal calibration for a value of 5 ((3d, (3h, (3i, blue curve).*

The manuscript has been modified in the flowing way:

- Material & method:

The initial description of the synthetic dataset experiment:

"Synthetic dataset experiment: A sequence of T(= 100) images was created. To simulate respiratory motion typically encountered on mobile organs, a "ground truth" periodic translational displacement field Dgt(t) (maximum amplitude=7.5 pixels, step size=1.5 pixels) was synthesized. The signal-tonoise ratio (SNR) of 20 was chosen to simulate a realistic acquisition (typically between 5 and 20). To simulate the presence of a structure appearing transient, a rectangular structure was added in half of the images (corresponding to the images with the three largest displacement amplitudes of each cycle)."

has been changed to (see page 4)

"*Synthetic dataset experiment:*

Data creation: A sequence of T(= 30) images was created. To simulate respiratory motion typically encountered on mobile organs, a periodic (period=6 frames) geometric transformation composed by a 2D translation (T_x,T_y) and scaling (S_x,S_y) was synthesized (T_x = {0, 0.5, 1, 1.5, 1, 0.5}, T_y = {0, 2.5, 5, 7.5, 5, 2.5} pixels, S_x = S_y = {1, 1.03, 1.06, 1.09, 1.06, 1.03}). A signal-to-noise ratio (SNR_{dB}) of 1.3 was chosen to simulate a realistic acquisition (typically between 0.7 and 1.3). A structure appearing transiently in the lower part of the image was added in half of the images to simulate the effect potentially encountered with reduced FOV imaging. Quality assessment of the motion estimation: Since the real motion (D_{gt} = (u_{GT}, v_{GT})) and the estimated motion (D = (u, v)) are available for each pixel in such synthetic dataset experiment, the measures commonly reported in the optical flow community such as the endpoint error (EE) and the angular error (AE) of the flow [9], [26] were computed, with:

$$EE = \sqrt{(u - u_{GT})^2 + (v - v_{GT})^2} \qquad (11)$$

$$AE = \cos^{-1}\left(\frac{1 + u \cdot u_{GT} + v \cdot v_{GT}}{\sqrt{1 + u^2 + v^2}\sqrt{1 + u_{GT}^2 + v_{GT}^2}}\right) \qquad (12)$$

In addition, to provide additional information about the smoothness of the estimated motion field, the harmonic energy of the estimated flow [27] was computed."

- Results

In the result section, the initial figure 2 has been replaced by two new figures (Figure 2,3). The initial synthetic result description:

"The comparison between our method and the H&S algorithm was first realized on a synthetic dataset where the reference image and an image corresponding to a different position of the motion cycle are displayed respectively in Fig. 2a and b. The

appearing rectangle, located at the bottom of Fig. 2b, is expected to hamper the motion estimation at the bottom of the object (in the red ROI displayed in Fig. 2a). Therefore, the gold standard error was computed for each dynamic, over this ROI and the results are plotted in Fig. 2c. The H&S algorithm showed very poor performance on this area where the appearing rectangular structure biased the accuracy of the algorithm (three repetitive high values). On the other hand, the proposed CME approach remained stable over the time and provided a more accurate motion field."

has been changed to (See page 5)

"An example of synthetic images is shown in Fig. 2 where the reference image (Fig. 2a, containing the transient structure) and an image corresponding to the maximum synthesized motion (Fig. 2b, 4th image of the cycle) are displayed. The contours (interior and exterior) of the target obtained from the reference image are reported in dashed red curves. The H&S approach was employed to register the image in Fig. 2b using different α^2 *values. When using a small value (*α^2 *= 0:5), the registration is strongly influenced by the underlying structure leading to a severe registration artifact (yellow arrows in Fig. 2d). By using an optimally calibrated value (*α^2 *= 2:5, see Fig. 3a), the registration artifact was reduced but still present (yellow arrows in 2e). A higher value of* α^2*(= 10) is reported in Fig. 2f where the registration artifact introduced by the transient structure appeared substantially reduced due to a higher weight on the smoothness of the motion field. However, in this case a different registration artifact was observed due to the inability of the method to handle complex motion (here the scaling effect) as shown by the yellow arrows. The CME approach provided a reliable registration where the registered target perfectly matched the reference target contour. These results were confirmed in the plot of the averaged EE values over time (computed inside the target, between the two red dashed curves) in Fig. 2g. While the H&S approach periodically failed to recover the correct motion, the CME offered more robust performances over time.*

In order to better characterize the potential gain and limitations of the compared methods, their performance in terms of averaged error endpoint, harmonic energy and averaged angular error were precisely investigated as a function of the employed parameters (Fig. 3). As previously observed in Fig. 2d, low α^2 *values (Fig. 3a, 3e and 3i) provided poor performance since the methods became very sensitive to the presence of the intrusive structure. On the contrary, high values limited the ability of the algorithm to estimate complex motion and also deteriorated the motion estimates as confirmed by the convergence of the harmonic energy toward a very small value. The averaged EE values obtained with an optimal* α^2 *calibration were 0.42 for the H&S approach (*α^2 *= 2:5) and 0.21 for the CME method (*α^2*=6.5) showing a reduction of the averaged EE by a factor of 2. The* λ^2 *value influence was then investigated ((Fig. 3b, 3f and 3j). As expected, small* λ^2 *values tend to the H&S performance and high values tend to the extrapolation of the constraint point motions (and their associated uncertainty) leading in both cases to a deterioration of the registration. A*

good calibration of λ^2 (=0.1, blue curve) provided a significant improvement of the motion estimation (Fig. 3b, 3j). The influence of the constraint point number (N) was then evaluated (Fig. 3c, 3g and 3k). Although an optimal value was reached around 20 points (Fig. 3c, 3k, blue curve), the sensitivity of the CME to this parameter was limited. Finally the influence of the bandwidth (R^2) provided an optimal calibration for a value of 5 (Fig. 3d, 3h, 3i, blue curve).”

- Discussion

While these results were not shown in the initial submission of the paper, a discussion had been already integrated to the manuscript. Therefore, this part was conserved in the revised version of the paper.

3.6.2 Algorithm calibration for vivo experiments:

The authors would like to apologize again for the lack of information about the in-vivo calibration of the employed algorithms in the first submission of the manuscript.

The employed parameters for the in-vivo study were initially calibrated based on the registration performance obtained on similar data obtained without reduced field of view imaging. The authors agree that some clarification are needed on this point and should be added to the manuscript.

The parameters calibration of a motion estimation algorithm is difficult in-vivo since the real motion is unknown. Especially, by considering the initial synthetic case, as correctly pointed out by, a higher α^2 value will increase the smoothness constraint smoothness of the motion and will reduce the registration artifact introduced by the intrusive structures. However, such configuration will also decrease the ability of the algorithm to estimate complex motion. We recently showed in

S. Roujol, M. Ries, C. Moonen, and S. Denis, “Automatic nonrigid calibration of image registration for real time mr-guided hifu ablations of mobile organs.” IEEE transactions on medical imaging, vol. 30, no. 10, p. 1737, 2011.

that an optimal range of α^2 values for the correction of the respiratory motion on abdominal imaging was around [0.1-0.5] and a plateau was generally reached within this range of values. We also pointed out in the discussion that using the lower range of such plateau may help to improve the robustness of the algorithm against motion not encountered during the calibration step (with a potential higher amplitude, due to a deeper breath for example).

However, the choice of an optimal α^2 value is the presented reduced FOV imaging case appears more tedious. The optimal α^2 value (0.1-05) obtained from full FOV imaging will provide an optimal registration in the entire target except in the area closed to the intrusive structure. On the other hand, a higher value of α^2 will decrease

this artifact but will deteriorate the registration in areas subject to complex deformations (which for the case of the heart may represent the whole myocardium due to its complex deformation). Therefore, the optimal calibration depends on the criterion to be optimized. Therefore, we believe that in term of clinical perspective it is more valuable to provide an algorithm working correctly in the majority of the organ (using a smaller α^2 value calibrated on full FOV images, but potentially leading to strong local registration artifact) than a higher value which might decrease the overall DICE value while affecting/degrading the majority of the registration (due to its inability to cope with complex motion).

A paragraph has been added to the material and method part to justify this point as follows (See page 4)

"In-vivo calibration of the employed algorithms is complex since it depends on the criterion to be optimized. The calibration of the H&S method (α^2 value) has to deal with contradictory effects. A small regularization of the motion field is required to enable the estimation of complex motion and to have a globally reliable estimated motion. On the other hand, a high constraint on the motion field smoothness would reduce the registration artifact induced by intrusive structures but at the same time will limit the ability of the algorithm to estimate complex motion. This may thus deteriorate the registration in the entire organ. Therefore, we calibrated the employed algorithms in a way to maintain optimal performance in the general case of full FOV imaging (without the presence of intrusive structures). As recently shown [24], a reliable in vivo calibration of the H&S algorithm was obtained for a range of α^2 values between 0.1-0.5. A plateau was generally observed for these ranges of values and its lower bound was suggested as a good way to cope with variations of the breathing pattern (such as an amplitude variation or drift of the respiration pattern). Therefore, a α^2 value of 0.1 was employed for the H&S algorithm. We empirically found a near-optimal solution for the CME calibration by employing the following parameters: α^2=0.1, λ^2=0.1, N=20 and R^2=5. Note that a similar optimal configuration was obtained for the synthetic dataset experiment (see Result section) except for the α^2 values that were higher due the lowest complexity of the synthesized motion."

Also as suggested by the reviewer, the harmonic energy was computed for all in-vivo cases and is now reported in Figure 4. A sentence has been added to the manuscript in the material and method section as follows (see page 5)

"The harmonic energy was also reported for all tested cases to assess the energy of the deformation fields."

In the result section the following sentences were added (See page 6)

"These findings were typical for the entire sequences as confirmed by the time evolutions of the DICE similarity coefficient and the harmonic energy, respectively

shown in Fig. 4e, 4m and Fig. 4f, 4n. Due to the respiratory cycle, the transient structures appeared periodically in the time series and the H&S approach periodically failed to recover a reliable motion estimate leading to low DICE similarity coefficient and elevated harmonic energy values."

…

"As expected, higher harmonic energy values were obtained with the H&S method since the method periodically failed and employed a relative lower weight of the smoothness constraint term (the same α2 value (=0.1) was employed for both algorithms)."

Finally the abstract was modified to integrate and emphasize the results obtained with the presented CME as follows (see page 1)

"Compared to the H&S approach, a significant improvement ($p<0.05$) of the DICE similarity criterion computed between the reference and the registered organ positions was achieved."

3.7 Lastly, regarding the synthetic experiment it would be interesting to show the angular error as typically done in the optical flow community.

The angular error has been computed and reported the new synthetic dataset. Since this point has been extensively discussed in the response 3.6.1, the authors invite the reviewer to refer to this answer.