Inverse Problems and Imaging (to appear 2018) FINAL DRAFT
(c) *American Institute of Mathematical Sciences //*
http://www.aimsciences.org/journal/1930-8337

An Augmented Lagrangian Method for Solving a New Variational Model based on Gradients Similarity Measures and High Order Regularization for Multimodality Registration *

Anis Theljani and Ke Chen†

Abstract. In this work we propose a variational model for multi-modal image registration. It minimizes a new functional based on using reformulated normalized gradients of the images as the fidelity term and higher-order derivatives as the regularizer. We first present a theoretical analysis of the proposed model. Then, to solve the model numerically, we use an augmented Lagrangian method (ALM) to reformulate it to a few more amenable subproblems (each giving rise to an Euler-Lagrange equation that is discretized by finite difference methods) and solve iteratively the main linear systems by the fast Fourier transform; a multilevel technique is employed to speed up the initialisation and avoid likely local minima of the underlying functional. Finally we show the convergence of the ALM solver and give numerical results of the new approach. Comparisons with some existing methods are presented to illustrate its effectiveness and advantages.

Key words. Variational model; Optimization; Multi-modality images; Similarity measures; Mapping; High order regularisation; Inverse Problem; Augmented Lagrangian; Multilevel.

AMS subject classifications.

1. Introduction.

Image registration consists in finding a reasonable spatial geometric transformation between given two images of the same object taken at different times or acquired using different devices. It is a challenging task required in diverse fields of astronomy, optics, biology, chemistry, medical imaging and remote sensing and particularly in medical imaging. For an overview of image registration methodology and approaches, we refer to [11, 23, 24, 31]. Here, we focus on deformable image registration for multi-modality images using variational approaches which belong to the class of the widely used methods ([2, 5, 7, 14, 21, 22, 36]) and aim to find a better gradients-based model than the standard gradient models.

It is informative to illustrate the notation of the image registration modelling by considering a pair of mono-modal images: Given a fixed image (also called reference) and a moving image (also called template), which are represented by scalar functions $T, R : \Omega \subset \mathbb{R}^d \longrightarrow \mathbb{R}$, find a reasonable geometric transformation $\varphi(\mathbf{u})(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, $\mathbf{u} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that:

$$T[\varphi(\mathbf{u})] = T(\mathbf{x} + \mathbf{u}(\mathbf{x})) = R. \tag{1.1}$$

This is an equation of the unknown displacement field $\mathbf{u}$, which is supposed to be sought in a properly chosen functional space. The reconstruction model (1.1) is an ill-posed inverse problem and thus regularisation techniques are needed to overcome ill-posedness. Generally, the regularisation technique turns an ill-posed problem such as model (1.1) into a well-posed one which minimizes an energy compromised of a regularisation term (mostly a semi-norm of

*Liverpool Centre for Mathematics in Healthcare and Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK. This work was funded by the UK EPSRC grant EP/N014499/1.
†Corresponding author. k.chen@liverpool.ac.uk. Web: www.liv.ac.uk/∼cmchenke

a functional space that is fixed a priori) and a data fidelity term. In summary, the desired displacement $\mathbf{u}$, in some appropriate space $\mathcal{H}$, is a minimizer of the following joint energy functional:

$$\min_{\mathbf{u}\in\mathcal{H}}\{\mathcal{J}(\mathbf{u}) = S(\mathbf{u}) + \frac{\lambda}{2}D(T(\mathbf{u}), R)\}. \tag{1.2}$$

This model may be used for registering both mono-modality and multi-modality images.

Here in (1.2), the first term $S(\mathbf{u})$ is a regularisation term which controls the smoothness of $\mathbf{u}$ and reflects our expectations by penalising unlikely transformations. Many works tackled the question of how to choose the best regularisation term that gives the more possible plausible transformation. Various regularizers have been proposed, such as first-order derivatives based on total variation [4, 16], diffusion [9] and elastic regularizer registration models and higher-order derivatives-based on linear curvature [10], mean curvature [6] and Gaussian curvature [17] models; we can refer also to [5, 22, 39, 40, 41].

The second term $D(T(\mathbf{u}), R)$ is a fidelity measure, which quantifies distance or similarity of the transformed template image $T(\mathbf{u})$ and the reference R, whereas λ is a positive weight which controls the trade-off between them. In the case of mono-modal images, the fixed and the moving images have similar features and same intensity ranges. Thus, either the L^1- distance (Sum of Absolute Differences) $D = \|T - R\|_1$ or the well-known choice L^2- distance (Sum of Squared Differences) between R and $T(\mathbf{u})$ i.e. $D = \|T - R\|_2^2 = \int_\Omega (T(\mathbf{u}) - R)^2 d\mathbf{x}$ may be used as a similarity measure. Clearly such a measure only makes sense for mono-modal images.

For a pair of multi-modal images T, R (generated from independent imaging techniques), unfortunately, one cannot minimize $\|T - R\|$ since values of T, R are not directly comparable. That is, only the patterns of T, R bear some resemblance to each other, not their values (so called intensity values). Therefore, intensities of the same object in different images are not similar which makes the problem much harder than the mono-modality case. Hence many good models as from [23] for mono-modal images and also the elegant mathematical approach of optimal transport [8] cannot be used. For multi-modal images, varous similarity measures have been used and include Mutual Information [20, 26, 33] and Normalised Gradient Field [15, 18, 30]. Recently [3] proposed a cross-correlation similarity measure based on reproducing kernel Hilbert spaces and found advantages over Mutual Information. Below, we briefly review these two commonly used measures: mutual information and normalized gradient fields.

Mutual Information (MI). It takes its origin from the theory of information and was firstly proposed in [33]. Several variants of MI approach were proposed in recent years (see [20, 26]), showcasing its great capability as well as limitations. The basic idea behind MI is the comparison of the histograms of the two images instead of comparing their intensities. The Mutual information between the two images if given by the following quantity:

$$D^{MI}(T(\mathbf{u}), R) = -\int_{\mathbb{R}^2} p_{T,R}(t, r) \log \frac{p_{T,R}(t, r)}{p_T(t)p_R(r)}\, dtdr, \tag{1.3}$$

where p_R, p_R are probability distributions of the gray values in R and T, whereas $p_{T,R}$ is the joint probability of the gray values which can be derived from the joint histogram. As the MI measure involves histograms, its inherent disadvantages are how to choose the size of bins and how to remedy the lack of spatial relationships to avoid mis-registrations. In addition, the measure also fails when features with different intensities in the first image have similar intensities in the second one [19], which is the case in perfusion imaging.

Normalised Gradient Field (NGF). The basic idea of the Normalised Gradient Field (NGF) [15, 18, 30] is the use of a derived information from the image intensity, i.e, the gradient. Similarity measures depending in the gradients or geometry of the images, which naturally encode information about the shape, can be better. The key idea behind the NGF measure is to align the gradients $\nabla T(\mathbf{u})$ and ∇R by minimizing the cosines distance between them. More precisely, on each point $x \in \Omega$, try to find a displacement $\mathbf{u}(x)$ such that $\cos\Theta = 1$ where Θ is the angle between $\nabla T(x + \mathbf{u}(x))$ and $\nabla R(x)$. Therefore, the NGF consists in minimization of the following energy:

$$D^{NGF}(T(\mathbf{u}), R) = \int_\Omega (1 - (\cos\Theta)^2)\, d\mathbf{x} = \int_\Omega (1 - (\nabla_n T(\mathbf{u}) \cdot \nabla_n R)^2)\, d\mathbf{x}, \tag{1.4}$$

where $\nabla_n T(\mathbf{u}) = \nabla T(\mathbf{u})/\|\nabla T(\mathbf{u})\|$ and $\nabla_n R = \nabla R/\|\nabla R\|$ are normalised unit vectors. As the NGF uses the product scalar between the two vectors $\nabla_n R$ and $\nabla_n T(\mathbf{u})$, it will not work well when the gradients are null or very weak. In other words, suppose that in a large region of the image T, we have $\nabla T \perp \nabla R$ and then $1 - (\nabla_n T \cdot \nabla_n R)^2) \approx 1$, which means that solving the optimization problem (1.2) is equivalent to only smoothing the deformation $\mathbf{u}$ in this region whereas the similarity measure does not play a role in the energy, which is not reliable. As an example, we consider the images in the Fig 1(a-b) where $\nabla_n T \cdot \nabla_n R = 0$ a.e in Ω due to one of $\nabla_n T$, $\nabla_n R$ being zero, we see that if we use the NGF in (1.2), there is no change in the template image because of the reason mentioned before, so $T(\mathbf{u})$, obtained using the NGF as measure, shown in Fig 1(c) is not correct. If we use the ratio #N of the number of pixels where $\nabla_n T \cdot \nabla_n R \neq 0$ over the total number of pixels, we have observed the current NGF would not give a good registration result if $\#N \leq 25\%$. In this work, believing in the elegance of geometric fitting, we aim to improve the above NGF for these cases. we are primarily motivated to explore the potential of normalised gradients beyond its standard form. Our question is whether or not a better normalised gradients-based model than the well-known form [15, 18, 30] is possible.

The outline of the paper is as follows. In Section 2, we propose our variational model which minimizes an energy with new similarity measures and we prove by variational techniques the existence of a minimizer. Section 3 is dedicated to the numerical solution of the proposed model by an augmented Lagrangian approach and analysis of convergence. Finally, Section 4 concerns the implementation and the presentation of several numerical examples to test the efficiency and robustness of the proposed approach.

2. The new multi-modality model.

Since our formulation consists of two building blocks: a similarity measure D and a regularization term S, we now discuss our choice of regularizers and the distance measure. Because our emphasis is on the latter, almost all regularizers suitable for variational registration models of mono-modal images may also be used.

Choice of Similarity Measure. To motivate our proposed measure D, consider the the NGF example in Fig 1. For this specific example, note that where $\nabla_n T \cdot \nabla_n R = 0$ we have

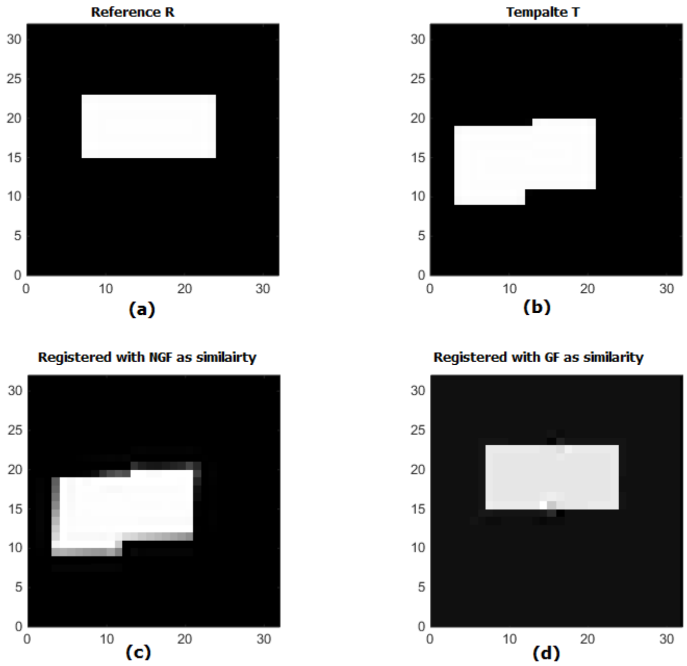


Figure 1. *Example of Reference and Template images where $\nabla_n T \cdot \nabla_n R = 0$ (or one of $\nabla_n T$, $\nabla_n R$ is zero) a.e in Ω.*

$\|\nabla_n T - \nabla_n R\| \neq 0$. This suggests a revised NGF model

$$\min_{\mathbf{u}} \{S(\mathbf{u}) + \frac{\lambda}{2} D^{GF}(T(\mathbf{u}), R)\},$$

with the new measure D^{GF} replacing the standard NGF measure D^{NGF}:

$$D^{GF}(T(\mathbf{u}), R) = \int_\Omega \mathbf{GF}(T(\mathbf{u}), R) d\mathbf{x}, \quad \text{where} \quad \mathbf{GF}(T(\mathbf{u}), R) = |\nabla_n T - \nabla_n R|^2.$$

As expected, such a model can solve the example from Fig 1(a-b) with acceptable registration result shown in Fig 1(d). This suggests that a better choice of normalised gradients as similarity measure is possible for multi-modal registration scenario. Moreover, to enhance this idea, we use Fig. 2 to show that alignment of two vectors $X = \nabla T, Y = \nabla R$ from a large discrepancy on the left to the small discrepancy on the right amounts to minimization of the distance $|X| + |Y| - |X + Y|$ (which is similar to minimizing $\cos\theta(X, Y)$ as in D^{NGF}). Below we shall combine the ideas of minimizing both $|X - Y|$ and $|X| + |Y| - |X + Y|$.

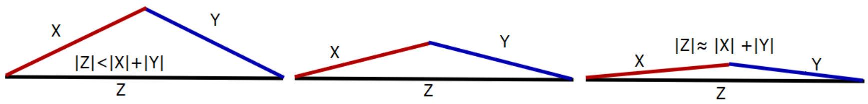


Figure 2. *Three examples of the triangle inequality for triangles with sides X, Y and Z. The left example shows a case where $|Z|$ is much less than the sum $|X| + |Y|$ of the other two sides, and the right example shows a case where $|Z|$ is only slightly less than $|X| + |Y|$.*

Choice of a Regularizer. As mentioned, there is a large class of possible regularizers that we could choose from. Here we choose a robust regulariser that allows large and smooth deformation, comprised of both first order and second order derivatives for the deformation field.

Based on the new measure, we propose to register the two functions R, T from different image modalities by solving the following minimization problem:

$$\begin{cases} \min\limits_{\mathbf{u}\in\mathcal{W}}\{\mathcal{J}_1(\mathbf{u}) = S(\mathbf{u}) + \frac{\lambda}{2}D^{GF}(T(\mathbf{u}), R) + \frac{\lambda}{2}D^{TM}(T(\mathbf{u}), R)\}, \\ \text{w.r.t} \quad \mathcal{C}(\mathbf{u}) = \det\left(I + \nabla\mathbf{u}\right) > 0, \end{cases} \tag{2.1}$$

where $\mathcal{W} = W_0^{1,2}(\Omega) \cap W^{2,2}(\Omega)$, $\mathcal{C}(\mathbf{u}) = (1 + \frac{\partial u_1}{\partial x}(1 + \frac{\partial u_2}{\partial y}) - \frac{\partial u_1}{\partial y}\frac{\partial u_2}{\partial x}$ and

$$S(\mathbf{u}) = \frac{\alpha}{2}\int_\Omega |\nabla\mathbf{u}|^2 d\mathbf{x} + \frac{\alpha_1}{2}\int_\Omega |\nabla^2\mathbf{u}|^2 d\mathbf{x}, \tag{2.2}$$

$$D^{GF}(T(\mathbf{u}), R) = \int_\Omega \mathbf{GF}(T(\mathbf{u}), R) d\mathbf{x}, \tag{2.3}$$

$$D^{TM}(T, R) = \int_\Omega \mathbf{TM}(T(\mathbf{u}), R)\, d\mathbf{x}. \tag{2.4}$$

where $\mathbf{TM}(T(\mathbf{u}), R) = (|\nabla T(\mathbf{u})| + |\nabla R| - |\nabla T(\mathbf{u}) + \nabla R|)^2$. Here in the term D^{GF}, we must use the normalized gradients rather than the usual gradients because the difference in the magnitude of gradients of R and $T(\mathbf{u})$ is large in multi-modality images. Moreover, we can easily prove that minimizing the length of $\mathbf{TM}(T(\mathbf{u}), R) = |\nabla T(\mathbf{u})| + |\nabla R| - |\nabla T(\mathbf{u}) + \nabla R|$ is equivalent to minimize the angle θ between the vectors $\nabla T(\mathbf{u})$ and ∇R, which leads to the alignment of the edges of R and $T(\mathbf{u})$; note that an alternative to minimizing the above $\mathbf{TM}$ is to minimize $\mathbf{TM}_n(T(\mathbf{u}), R) = |\nabla_n T(\mathbf{u})| + |\nabla_n R| - |\nabla_n T(\mathbf{u}) + \nabla_n R|$ based on normalized gradients. However, this will lead to a more difficult problem to solve numerically due to higher non-linearity. Our primary choice for regularization is the diffusion model [9] which uses first-order derivatives and promotes smoothness. As affine linear transformations are not included in the kernel of the H^1-regularizer, we desire a regularizer which can penalize such transformation. As such, we add the regularizer based on second-order derivatives (LLT) to the model which allows to remove the need of any preregistration step of affine transformation. The second-order derivatives allows also getting smooth transformations [41]. The constraint $\mathcal{C}(\mathbf{u}) > 0$ on the determinant in the minimization problem (2.1) guarantees that the resulting deformation field $\varphi = \mathbf{x} + \mathbf{u}$ suffers no mesh folding and thus is physically plausible; see also [12, 13, 28]. Different alternatives were proposed to ensure invertibility by adding another regularisation term depending on the determinant of the transformation to the registration objective function; see [2].

Mathematical analysis of the proposed model. Most registration models are non-convex with respect to $\mathbf{u}$ and consequently, if solutions exist, there are local minimizers or solutions are generally not unique. Below we prove the existence of a minimizer for problem (2.1). Before stating the main result, we first consider the concept of Carathéodory functions.

Definition 2.1. *Let $\Omega \subset \mathbb{R}^d$ be an open set and let $f : \Omega \times \mathbb{R}^n \times \mathbb{R}^{d\times n} \times \mathbb{R}^{d\times d\times n} \to [0, +\infty)$. Then f is a Carathéodory function if:*
1. $f(x, \cdot, \cdot, \cdot)$ is continuous for almost every $x \in \Omega$.
2. $f(x, \mathbf{u}, \psi, \Theta)$ is measurable in x for every $(\mathbf{u}, \psi, \Theta) \in \mathbb{R}^n \times \mathbb{R}^{d\times n} \times \mathbb{R}^{d\times d\times n}$.

We will use some theory about integrals of higher-order. It also sets up assumptions with which our optimisation problem (2.1) admits a minimiser.

Lemma 2.2 ([42]). *Let $\Omega \subset \mathbb{R}^d$ be an open set and $f : \Omega \times \mathbb{R}^n \times \mathbb{R}^{d\times n} \times \mathbb{R}^{d\times d\times n} \to [0, +\infty)$ satisfies the following assumptions:*
(i) f is a Carathéodory function.
(ii) $f(x, \mathbf{u}, \psi, \Theta)$ is quasi-convex with respect to Θ.
(iii) $0 \le f(x, \mathbf{u}, \psi, \Theta) \le a(x) + C(|\psi|^p + |\Theta|^p)$ where $a(x) \in L^1(\Omega)$, $C > 0$.
Then $\mathcal{J}(\mathbf{u})$ is weak lower semi-continuous (denoted by wlsc) in $\mathcal{W}$.

To analyse the proposed model (2.1), it is convenient to rewrite the energy $\mathcal{J}(\cdot)$ by merging all terms under one integral in the following form:

$$\mathcal{J}(\mathbf{u}) = \int_\Omega f(x, \mathbf{u}, \nabla\mathbf{u}, \nabla^2\mathbf{u})\, d\mathbf{x},$$

where
$$f(x, \mathbf{u}, \psi, \Theta) = \frac{\alpha}{2}|\psi|^2 + \frac{\alpha_1}{2}|\Theta|^2 + \frac{\lambda}{2}|\nabla_n T(\mathbf{u}) - \nabla_n R|^2 + \frac{\lambda}{2}(|\nabla T(\mathbf{u})| + |\nabla R| - |\nabla T(\mathbf{u}) + \nabla R|)^2,$$
To apply the Lemma 2.2, we assume that $|\nabla R|$ and $|\nabla T(\mathbf{u})|$ are bounded almost everywhere by a constant $c > 0$. Then, we have the following result:

Lemma 2.3. *The energy functional $\mathcal{J}(\cdot)$ is coercive and wlsc in $\mathcal{W}$.*

Proof. The coercivity can easy obtained using the Poincaré inequality. In fact, the later guarantees that

$$\|\mathbf{u}\|_{\mathcal{W}} = \left(\|\nabla\mathbf{u}\|_2^2 + \|\nabla^2\mathbf{u}\|_2^2\right)^{1/2}$$

defines a norm in the space $\mathcal{W}$. Using the positivity of $D^{GF}(T(\mathbf{u}), R)$ and $D^{TM}(T(\mathbf{u}), R)$, we have:

$$\mathcal{J}(\mathbf{u}) \ge \frac{\min(\alpha, \alpha_1)}{2}\|\mathbf{u}\|_{\mathcal{W}}^2,$$

which directly gives the coercivity of $\mathcal{J}(\cdot)$. For the weak lower semi-continuity, we now verify that the functions $f(\cdot)$ fulfils the assumptions in Lemma 2.2:
i) Since the gradient of the fixed and the moving image ∇R and $\nabla T(\mathbf{u})$ are assumed to be continuous, $f(\cdot)$ is Carathéodory function.
ii) It is easy to check that $f(x, \mathbf{u}, \psi, \Theta)$ are convex with respect to Θ, clearly implying that it is quasi-convex.
iii) For condition (iii), we have $|\nabla_n T(\mathbf{u})| \le 1$ and $|\nabla_n R| \le 1$, which means that:

$$\frac{\lambda}{2}|\nabla_n T(\mathbf{u}) - \nabla_n R|^2 \le \frac{\lambda}{2}(|\nabla_n T(\mathbf{u})| + |\nabla_n R|)^2 \le 2\lambda. \tag{2.5}$$

Moreover, using the fact that $|\nabla R|$ and $|\nabla T(\mathbf{u})|$ are bounded almost everywhere by a constant $c > 0$, we get

$$\frac{\lambda}{2}(|\nabla T(\mathbf{u})| + |\nabla R| - |\nabla T(\mathbf{u}) + \nabla R|)^2 \leq \frac{\lambda}{2}(|\nabla T(\mathbf{u})| + |\nabla R|)^2 \leq 2\lambda c^2. \tag{2.6}$$

Therefore, using inequalities (2.5) and (2.6), we have:

$$\begin{aligned} f(x, \mathbf{u}, \psi, \Theta) &= \frac{\alpha}{2}|\psi|^2 + \frac{\alpha_1}{2}|\Theta|^2 + \frac{\lambda}{2}|\nabla_n T(\mathbf{u}) - \nabla_n R|^2 \\ &\qquad + \frac{\lambda}{2}(|\nabla T(\mathbf{u})| + |\nabla R| - |\nabla T(\mathbf{u}) + \nabla R|)^2 \\ &\leq \frac{\alpha}{2}|\psi|^2 + \frac{\alpha_1}{2}|\Theta|^2 + 2\lambda c^2 + 2\lambda. \end{aligned}$$

Then, the function $f(\cdot)$ fulfils the condition (iii) of Lemma 2.2 with $a(x) \equiv \lambda c^2 + 2\lambda$ which implies that the energy $\mathcal{J}(\cdot)$, is *wlsc* in $\mathcal{W}$. ∎

We are now ready to prove the existence of a solution for the minimization model (2.1). Based on Lemma 2.2 and Lemma 2.3, we have the following result:

Proposition 2.4. *The minimization problem (2.1) admits at least one solution in the space $\mathcal{A} = \{\mathbf{u} \in \mathcal{W}; \mathcal{C}_\epsilon(\mathbf{u}) \geq 0\}$ where $\epsilon > 0$ is a small parameter, $\mathcal{C}_\epsilon(\mathbf{u}) = \mathcal{C}(\mathbf{u}) - \epsilon$, and $\mathcal{C}(\cdot)$ is given in (2.1).*

Proof. Consider a minimizing sequence $(\mathbf{u}_n)_n \subset \mathcal{A}$ of $\mathcal{J}(\cdot)$, i.e.,

$$\mathcal{J}(\mathbf{u}_n) \underset{n\to\infty}{\longrightarrow} \inf_{\mathbf{u}\in\mathcal{A}} \mathcal{J}(\mathbf{u}).$$

The coercivity of $\mathcal{J}(\cdot)$ guarantees that the sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}$ is uniformly bounded $\mathcal{W}$. Thus, there exists a subsequence, still denoted $(\mathbf{u}_n)_{n\in\mathbb{N}}$, such that $\mathbf{u}_n \underset{n\to\infty}{\rightharpoonup} \mathbf{u}$ weakly in $\mathcal{W}$. Using the weak lower semi-continuity of $\mathcal{J}(\cdot)$, we obtain that the limit $\mathbf{u}$ is a minimizer of $\mathcal{J}(\cdot)$. It remains to prove that $\mathbf{u}$ fulfils the constraint $\mathcal{C}(\mathbf{u}) > 0$. Now, we show that $\mathcal{A}$ is weakly closed subset of $\mathcal{W}$. Let $\mathbf{u}_k$ be a weakly convergent sequence to $\mathbf{u}$ in $\mathcal{W}$. From the definition of the space $\mathcal{W}$, we have that $\mathbf{u}_k$ is weakly convergent to $\mathbf{u}$ in $W^{1,2}(\Omega)$ and $\mathbf{u}_k$ is weakly convergent to $\mathbf{u}$ in $W^{2,2}(\Omega)$. Moreover, as the sets $\mathcal{A}_1 = \{\mathbf{u} \in W^{1,2}(\Omega); \mathcal{C}_\epsilon(\mathbf{u}) \geq 0\}$ and $\mathcal{A}_2 = \{\mathbf{u} \in W^{2,2}(\Omega); \mathcal{C}_\epsilon(\mathbf{u}) \geq 0\}$ are weakly closed for $W^{1,2}$-topology and $W^{1,2}$-topology (see [27]), respectively, we get that $u \in \mathcal{A}_1$ and $u \in \mathcal{A}_2$. Then, the limit $\mathbf{u}$ belongs to the intersection $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ and thus $\mathcal{A}$ is weakly closed. Therefore, the minimizer $\mathbf{u}$ belongs to the set $\mathcal{A}$, i.e., $\mathcal{C}(\mathbf{u}) \geq \epsilon > 0$, which finishes the proof. ∎

3. Augmented Lagrangian method (ALM).

The energies $\mathcal{J}(\cdot)$ are highly non-linear, and their numerical resolution is a non-trivial task. Thus, we propose an Augmented Lagrangian Method (ALM) which is often used for solving constrained minimization problems by replacing the original problem by an unconstrained problem. The method is similar to the penalty method where the constraints are incorporated in the objective functional and the problem is solved using alternating minimization of the sub-problems; see [1, 29, 32, 35, 43, 44] for various successful applications.

3.1. ALM iterations. Introducing three intermediate variables K, $\mathbf{p}$ and $\mathbf{n}$ to reformulate (2.1), we solve the following constrained minimization problem:

$$
\begin{cases} \min\limits_{\mathbf{u},K,\mathbf{p},\mathbf{n}} \{S(\mathbf{u}) + \frac{\lambda}{2}\int_\Omega (\mathbf{n} - \nabla_n R)^2 d\mathbf{x} + \frac{\lambda}{2}\int_\Omega (|\mathbf{p}| + |\nabla R| - |\mathbf{m}|)^2\, d\mathbf{x}\}, \\ \text{w.r.t} \quad K = T(\mathbf{u}), \quad \mathbf{p} = \nabla K, \quad |\mathbf{p}|\mathbf{n} = \mathbf{p}, \quad \mathbf{m} = \mathbf{p} + \nabla R, \quad \mathcal{C}_\epsilon(\mathbf{u}) \geq 0. \end{cases} \tag{3.1}
$$

Then, the augmented Lagrangian functional corresponding to the constrained optimization problem (3.1) is defined as follows:

$$
\begin{aligned}
&\mathcal{L}_1(\mathbf{u}, K, \mathbf{p}, \mathbf{n}, \mathbf{m}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \\
&= S(\mathbf{u}) + \frac{\lambda}{2}\int_\Omega (\mathbf{n} - \nabla_n R)^2 d\mathbf{x} + \frac{\lambda}{2}\int_\Omega (|\mathbf{p}| + |\nabla R| - |\mathbf{m}|)^2\, d\mathbf{x} \\
&\quad + \frac{r_2}{2}\int_\Omega (\mathbf{p} - \nabla K)^2 d\mathbf{x} + \frac{r_3}{2}\int_\Omega (\mathbf{p} - |\mathbf{p}|\mathbf{n})^2 d\mathbf{x} + \frac{r_4}{2}\int_\Omega (\mathbf{p} + \nabla R - \mathbf{m})^2 d\mathbf{x} \\
&\quad + \int_\Omega (T(\mathbf{u}) - K)\lambda_1 d\mathbf{x} + \int_\Omega (\mathbf{p} - \nabla K)\cdot \lambda_2 d\mathbf{x} + \int_\Omega (\mathbf{p} - |\mathbf{p}|\mathbf{n}) \cdot \lambda_3 d\mathbf{x} \\
&\quad + \int_\Omega (\mathbf{p} + \nabla R - \mathbf{m}) \cdot \lambda_4\, d\mathbf{x} + \frac{r_1}{2}\int_\Omega (T(\mathbf{u}) - K)^2 d\mathbf{x} + \frac{1}{2\sigma}\int_\Omega \mathcal{C}_s(\mathbf{u}, \lambda_5)\, d\mathbf{x},
\end{aligned} \tag{3.2}
$$

where

$$
\mathcal{C}_s(\mathbf{u}, \lambda_5) = [\min\{0, \sigma \mathcal{C}_\epsilon(\mathbf{u}) - \lambda_5\})]^2 - \lambda_5^2, \tag{3.3}
$$

$\sigma > 0$ and $\lambda_i, (i = 1, \cdots, 5)$ are the Lagrange multipliers. Since the optimisation problem (2.1) admits a minimizer, the previous augmented Lagrangian admits a saddle point $(\mathbf{u}^*, K^*, \mathbf{p}^*, \mathbf{n}^*, \mathbf{m}^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*)$.

3.2. Discretization and sub-problems. The images and the displacement fields are discretized on a uniform mesh using vertex centred discretization. We assume that the discrete solution $\mathbf{u}_{i,j} = \mathbf{u}(x_i, y_j)$, $i = 1, \cdots, l$, $j = 1, \cdots, c$ have $l \times c$ pixels, where l and c are the numbers of rows and columns in the image, respectively. Other quantities are set up similarly.

For sake of simplicity, we use a generic notation u for discussing discretization. For the discrete differential operators, we assume periodic boundary conditions for u. By choosing periodic boundary conditions, the action of each of the discrete differential operators can be regarded as a circular convolution of u and allows the use of fast Fourier transform (see [25, 34, 38] for more details). The discrete gradient is an operator from $\mathbb{R}^{l\times c}$ to $\mathbb{R}$, given by $\nabla u = (\partial_x u, \partial_y u)$ where ∂_x and ∂_y are *forward* difference operators defined as follows:

$$
\partial_x u = \begin{cases} u(i+1, j) - u(i, j) & 1 \leq i < l, 1 \leq j \leq c, \\ u(1, j) - u(i, j) & i = l, 1 \leq j \leq c, \end{cases}
$$

$$
\partial_y u = \begin{cases} u(i, j+1) - u(i, j) & 1 \leq i \leq l, 1 \leq j < c, \\ u(i, 1) - u(i, j) & 1 \leq i \leq l, j = c. \end{cases}
$$

The discrete divergence is an operator from $\mathbb{R}^{l\times c}$ to $\mathbb{R}$ and, for $\mathbf{n} = (n_1, n_2)$, given by $\operatorname{div}\mathbf{n} =$

$\overleftarrow{\partial}_x n_1 + \overleftarrow{\partial}_y n_2$ where backward difference operators are defined by

$$\overleftarrow{\partial}_x u = \begin{cases} u(i,j) - u(i-1,j) & 1 < i \leq l,\ 1 \leq j \leq c, \\ u(i,j) - u(l,j) & i = 1,\ 1 \leq j \leq c, \end{cases}$$

$$\overleftarrow{\partial}_y u = \begin{cases} u(i,j) - u(i,j-1) & 1 \leq i \leq l,\ 1 < j \leq c, \\ u(i,j) - u(i,c) & 1 \leq i \leq l,\ j = 1. \end{cases}$$

Then, the discrete Laplace operator is given by $\Delta u = \operatorname{div}(\nabla u)$. Similarly, we define the following (*forward* and *backward*) second-order discrete differential operators:

$$\partial_{xx} u = \overleftarrow{\partial}_{xx} u = \begin{cases} u(l,j) - 2u(i,j) + u(i+1,j) & i = 1, 1 \leq j \leq c, \\ u(i-1,j) - 2u(i,j) + u(i+1,j) & 1 < i < l, 1 \leq j \leq c, \\ u(i-1,j) - 2u(i,j) + u(1,i) & i = l, 1 \leq j \leq c, \end{cases}$$

$$\partial_{yy} u = \overleftarrow{\partial}_{yy} u = \begin{cases} u(i,c) - 2u(i,j) + u(i,j+1) & 1 \leq i \leq l,\ j = 1, \\ u(i,j-1) - 2u(i,j) + u(i,j+1) & 1 \leq i \leq l,\ 1 < j < c, \\ u(i,j-1) - 2u(i,j) + u(i,1) & 1 \leq i \leq l,\ j = c, \end{cases}$$

$$\partial_{xy} u = \partial_{yx} u = \begin{cases} u(i,j) - u(i+1,j) - u(i,j+1) + u(i+1,j+1) & 1 \leq i < l, 1 \leq j < c, \\ u(i,j) - u(1,j) - u(i,j+1) + u(1,j+1) & i = l, 1 \leq j < c, \\ u(i,j) - u(i+1,j) - u(i,1) + u(i+1,1) & 1 \leq i < l, j = c, \\ u(i,j) - u(1,j) - u(i,1) + u(1,1) & i = l, j = c, \end{cases}$$

$$\overleftarrow{\partial}_{xy} u = \overleftarrow{\partial}_{yx} u \begin{cases} u(i,j) - u(i,c) - u(l,j) + u(l,c) & i = l, j = 1, \\ u(i,j) - u(i,j-1) - u(l,j) + u(l,j-1) & i = 1, 1 \leq j < c, \\ u(i,j) - u(i,c) - u(i-1,j) + u(i-1,c) & 1 < i < l, j = 1, \\ u(i,j) - u(i,j-1) - u(i-1,j) + u(i-1,j-1) & 1 < i < l, 1 < j \leq c. \end{cases}$$

Based on the above operators, we define the following fourth-order differential operator:

$$\operatorname{div}^2.\nabla^2 u = \overleftarrow{\partial}_{xx}\partial_{xx} u + \overleftarrow{\partial}_{yy}\partial_{yy} u + \overleftarrow{\partial}_{xy}\partial_{xy} u + \overleftarrow{\partial}_{yx}\partial_{yx} u.$$

Thus the first version of an ALM algorithm is shown in Algorithm 3.1.

In order to solve the optimisation problem (3.4) more efficiently, we now consider a decoupled version of all main variables for the solution. The minimization problem is decomposed into a number of sub-problems, each of which can be solved quickly. In particular, we split the problem into four (main) sub-problems. Then, an alternating minimization and iterative procedure is obtained and shown in Algorithm 3.2. We discuss next how to solve these sub-problems.

The $\mathbf{u}$-subproblem. Fixing K^k, $\mathbf{p}^k$, $\mathbf{n}^k$, $\mathbf{m}^k$ and λ_i^k ($i = 1, \ldots, 5$), the $\mathbf{u}$-subproblem consists in finding $\mathbf{u}^{k+1}$ from solving the following minimization problem:

$$\min_{\mathbf{u}}\{S(\mathbf{u}) + \frac{r_1}{2}\int_\Omega (T(\mathbf{u}) - K^k)^2 d\mathbf{x} + \int_\Omega (T(\mathbf{u}) - K^k)\lambda_1^k d\mathbf{x} + \frac{1}{2\sigma}\int_\Omega \mathcal{C}_s(\mathbf{u}, \lambda_5^k)\, d\mathbf{x}\}. \tag{3.10}$$

Algorithm 3.1 Augmented Lagrangian method

1. Initialization: $\mathbf{u}^0$, K^0, $\mathbf{p}^0$, $\mathbf{n}^0$,$\mathbf{m}^0$ and λ_1^0, λ_2^0, λ_3^0, λ_4^0 and λ_5^0.
2. Iterate for $k = 1, 2, \ldots$ until a required tolerance:

— compute an approximate minimizers $\mathbf{u}^{k+1}$, K^{k+1}, $\mathbf{p}^{k+1}$, $\mathbf{n}^{k+1}$ and $\mathbf{m}^{k+1}$ of the augmented Lagrangian functional with the fixed Lagrange multipliers λ_1^k, λ_2^k, λ_3^k, λ_4^k and λ_5^k:

$$\begin{aligned}\left[\mathbf{u}^{k+1}, K^{k+1}, \mathbf{p}^{k+1}, \mathbf{n}^{k+1}, \mathbf{m}^{k+1}\right] = \\ \operatorname{argmin}_{\mathbf{u},K,\mathbf{p},\mathbf{n}} \mathcal{L}_1(u, K, \mathbf{p}, \mathbf{n}, \mathbf{m}, \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k, \lambda_5^k).\end{aligned} \tag{3.4}$$

— Update Lagrange multipliers

$$\lambda_1^{k+1} = \lambda_1^k + r_1(T(\mathbf{u}^{k+1}) - K^{k+1}), \tag{3.5}$$

$$\lambda_2^{k+1} = \lambda_2^k + r_2(\mathbf{p}^{k+1} - \nabla K^{k+1}), \tag{3.6}$$

$$\lambda_3^{k+1} = \lambda_3^k + r_3(\mathbf{p}^{k+1} - |\mathbf{p}^{k+1}|\mathbf{n}^{k+1}), \tag{3.7}$$

$$\lambda_4^{k+1} = \lambda_4^k + r_4(\mathbf{m}^{k+1} - \mathbf{p}^{k+1} - \nabla R), \tag{3.8}$$

$$\lambda_5^{k+1} = \max\{0, \lambda_5^k - \sigma\mathcal{C}_\epsilon(\mathbf{u}^{k+1})\}, \tag{3.9}$$

Algorithm 3.2 An more efficient solution procedure for alternating iterations

1. Initialization: $\tilde{\mathbf{u}}^0 = \mathbf{u}^k$, $\tilde{K}^0 = K^k$, $\tilde{\mathbf{p}}^0 = \mathbf{p}^k$,$\tilde{\mathbf{n}}^0 = \mathbf{n}^k$ and $\tilde{\mathbf{m}}^0 = \mathbf{m}^k$.
2. Iterate for $k = 1, 2, \ldots$ until a required tolerance:

— Set the Lagrange multipliers
$\lambda_1 = \lambda_1^k$, $\lambda_2 = \lambda_2^k$, $\lambda_3 = \lambda_3^k$, $\lambda_4 = \lambda_4^k$ and $\lambda_5 = \lambda_5^k$,

— Solve for $l = 1, \cdots, L$ the following problems:

$$\begin{aligned}\tilde{\mathbf{u}}^{l+1} &= \operatorname{argmin}_{\mathbf{u}} \mathcal{L}_1(\mathbf{u}, \tilde{K}^l, \tilde{\mathbf{p}}^l, \tilde{\mathbf{n}}^l, \mathbf{m}^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5),\\ \tilde{K}^{l+1} &= \operatorname{argmin}_{K} \mathcal{L}_1(\tilde{\mathbf{u}}^{l+1}, K, \tilde{\mathbf{p}}^l, \tilde{\mathbf{n}}^l, \mathbf{m}^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5),\\ \tilde{\mathbf{p}}^{l+1} &= \operatorname{argmin}_{\mathbf{p}} \mathcal{L}_1(\tilde{\mathbf{u}}^{l+1}, \tilde{K}^{l+1}, \mathbf{p}, \tilde{\mathbf{n}}^l, \mathbf{m}^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5),\\ \tilde{\mathbf{n}}^{l+1} &= \operatorname{argmin}_{\mathbf{n}} \mathcal{L}_1(\tilde{\mathbf{u}}^{l+1}, \tilde{K}^{l+1}, \tilde{\mathbf{p}}^{l+1}, \mathbf{n}, \mathbf{m}^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5),\\ \tilde{\mathbf{m}}^{l+1} &= \operatorname{argmin}_{\mathbf{n}} \mathcal{L}_1(\tilde{\mathbf{u}}^{l+1}, \tilde{K}^{l+1}, \tilde{\mathbf{p}}^{l+1}, \mathbf{n}^{l+1}, \mathbf{m}, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5).\end{aligned}$$

— Prepare for the next iteration by setting
$[\mathbf{u}^{k+1}, K^{k+1}, \mathbf{p}^{k+1}, \mathbf{n}^{k+1}, \mathbf{m}^{k+1}] = [\tilde{\mathbf{u}}^{l+1}, \tilde{K}^{l+1}, \tilde{\mathbf{p}}^{l+1}, \tilde{\mathbf{n}}^{l+1}, \tilde{\mathbf{m}}^{l+1}]$.

It is clear that the above minimization problem admits at least a solution $\mathbf{u} = (u_1, u_2)$ by solving the following system of PDEs in Ω:

$$\begin{cases} -\alpha\Delta u_1^{k+1} + \alpha_1 \mathrm{div}^2.\nabla^2 u_1^{k+1} + r_1(T(\mathbf{u}^{k+1}) - K^k)\partial_x T(\mathbf{u}^{k+1}) \\ \qquad + \lambda_1^k \partial_x T(\mathbf{u}^{k+1}) + \partial_{u_1}\mathcal{C}_s(\mathbf{u}^{k+1}, \lambda_5^k) = 0, \\ -\alpha\Delta u_2^{k+1} + \alpha_1 \mathrm{div}^2.\nabla^2 u_2^{k+1} + r_1(T(\mathbf{u}^{k+1}) - K^k)\partial_y T(\mathbf{u}^{k+1}) \\ \qquad + \lambda_1^k \partial_y T(\mathbf{u}^{k+1}) + \partial_{u_2}\mathcal{C}_s(\mathbf{u}^{k+1}, \lambda_5^k) = 0 \end{cases} \tag{3.11}$$

with the periodic boundary conditions on $\partial\Omega$. To solve the previous non-linear PDEs, we use a fast time marching method, i. e., find $\mathbf{u}^{k+1} = (u_1^{k+1}, u_2^{k+1})$ which solves

$$\begin{cases} u_1^{k+1} - dt[\alpha\Delta u_1^{k+1} + \alpha_1 \mathrm{div}^2.\nabla^2 u_1^{k+1}] = F_1(\mathbf{u}_{old}^{k+1}), & in\ \ \Omega, \\ u_2^{k+1} - dt[\alpha\Delta u_2^{k+1} + \alpha_1 \mathrm{div}^2.\nabla^2 u_2^{k+1}] = F_2(\mathbf{u}_{old}^{k+1}), & in\ \ \Omega, \end{cases} \tag{3.12}$$

where dt is the time step, $\mathbf{u}_{old}^{k+1}$ is the solution at the previous iteration for the time marching method and

$$F_1(\mathbf{u}) = -dt[r_1(T(\mathbf{u}) - R)\partial_x T(\mathbf{u}) - \lambda_1^k \partial_x T(\mathbf{u}) - \partial_{u_1}\mathcal{C}_s(\mathbf{u}, \lambda_5^k)] + u_1,$$

$$F_2(\mathbf{u}) = -dt[r_1(T(\mathbf{u}) - R)\partial_y T(\mathbf{u}) - \lambda_1^k \partial_y T(\mathbf{u}) - \partial_{u_2}\mathcal{C}_s(\mathbf{u}, \lambda_5^k)] + u_2.$$

To solve the above fourth-order equations in each time step iteration, we use the 2-dimensional discrete Fourier transforms. In fact, we have:

$$L \odot \mathcal{F}(u_1^{k+1}) = \mathcal{F}(F_1(\mathbf{u}_{old}^{k+1})), \text{ and } L \odot \mathcal{F}(u_2^{k+1}) = \mathcal{F}(F_2(\mathbf{u}_{old}^{k+1})),$$

where $L = I + \alpha dt\mathcal{F}(\Delta\cdot) + \alpha_1 dt\mathcal{F}(\mathrm{div}^2.\nabla^2\cdot)$. The operator $\mathcal{F}(\cdot)$ is the Fourier transform and "$\odot$" means point-wise multiplication of matrices. Therefore, the discrete solutions u_1 and u_2 can be obtained by applying the inverse of the discrete two-dimensional Fourier transform to the previous equation and we have:

$$u_1^{k+1} = \mathcal{F}^{-1}\left(\mathcal{F}(F_1(\mathbf{u}_{old}^{k+1})) \oslash L\right) \text{ and } u_2^{k+1} = \mathcal{F}^{-1}\left(\mathcal{F}(F_2(\mathbf{u}_{old}^{k+1})) \oslash L\right), \tag{3.13}$$

where "$\oslash$" means point-wise division of matrices.

Remark 1. *We emphasizes that computing the determinant is a non-trivial task. A discretization which well ensures that the map is diffeomorphic is discussed in [2, 13] and is based on finite element method. In our case, we are not using this discretization in the numerical computation as we are solving a system of PDEs defined only on the nodal points. However, the discretization is used for computing the determinant after getting the solution to check if the obtained map is diffeomorphic.*

The K-subproblem. Fixing $\mathbf{u}^{k+1}$, $\mathbf{p}^k$, $\mathbf{n}^k$, $\mathbf{m}^k$ and λ_i^k $(i = 1, \cdots, 5)$, the K-problem involves the minimization of the following energy:

$$\min_K \Big\{ \frac{r_1}{2}\int_\Omega (T(\mathbf{u}^{k+1}) - K)^2 d\mathbf{x} + \frac{r_2}{2}\int_\Omega (\mathbf{p}^k - \nabla K)^2 d\mathbf{x} + \int_\Omega (T(\mathbf{u}^{k+1}) - K)\lambda_1^{k-1} d\mathbf{x} + \int_\Omega (\mathbf{p}^k - \nabla K)\cdot \lambda_2^k d\mathbf{x} \Big\}.$$

This minimization problem is solved through its optimality condition:

$$-r_2\Delta K^{k+1} + r_1 K^{k+1} = r_1 T(\mathbf{u}^{k+1}) - r_2 \mathrm{div}\,\mathbf{p}^k - \mathrm{div}\,\lambda_2^k + \lambda_1^k. \tag{3.14}$$

We take advantage from the use of the 2-dimensional discrete Fourier transforms to compute K. In fact, applying the Fourier transforms to

$$LS \odot \mathcal{F}(K) = \mathcal{F}(RS),$$

where "$\odot$" means point-wise multiplication of matrices, RS is the right side of (3.14) and

$$LS = -r_2\mathcal{F}(\Delta\cdot) + r_1 I.$$

Therefore, the discrete solution is given by:

$$K = \mathcal{F}^{-1}\left(\mathcal{F}(RS) \oslash LS\right), \tag{3.15}$$

where $\mathcal{F}^{-1}(\cdot)$ is the inverse of the discrete two-dimensional Fourier transform.

The p-subproblem. Fixing $\mathbf{u}^{k+1}$, K^{k+1}, $\mathbf{n}^k$, $\mathbf{m}^k$ and λ_i^k $(i = 1, \cdots, 5)$, the $\mathbf{p}$-subproblem consists in minimizing, w.r.t., $\mathbf{p}$, the following energy:

$$\begin{aligned} &\frac{r_2}{2}\int_\Omega (\mathbf{p} - \nabla K^{k+1})^2 d\mathbf{x} + \frac{r_3}{2}\int_\Omega (\mathbf{p} - |\mathbf{p}|\mathbf{n}^k)^2 d\mathbf{x} + \frac{r_4}{2}\int_\Omega (\mathbf{p} + \nabla R - \mathbf{m}^k)^2 \\ &+ \int_\Omega (\mathbf{p} - \nabla K^{k+1}) \cdot \lambda_2^k d\mathbf{x} + \int_\Omega (\mathbf{p} - |\mathbf{p}|\mathbf{n}^k) \cdot \lambda_3^k + \int_\Omega (\mathbf{p} + \nabla R - \mathbf{m}^k) \cdot \lambda_4^k \, d\mathbf{x} \\ &+ \frac{\lambda}{2}\int_\Omega (|\mathbf{p}| + |\nabla R| - |\mathbf{m}|)^2 \, d\mathbf{x}. \end{aligned} \tag{3.16}$$

It is challenging to solve the above $\mathbf{p}$-minimization problem due to the non-differentiability of $|\mathbf{p}|$ in the quadratic term. To alleviate this situation, we consider a fixed-point formulation by lagging $|\mathbf{p}^k|\mathbf{n}^k$ in the k^{th} iteration instead of the constraint $\mathbf{p} = |\mathbf{p}|\mathbf{n}^k$. Thus, a simple reformulation rewrites the above problem as an equivalent minimization problem:

$$\min_{\mathbf{p}} \int_\Omega \beta|\mathbf{p}| \, d\mathbf{x} + \frac{r_2 + r_3 + r_4 + \lambda}{2}\int_\Omega (\mathbf{p} - C)^2 d\mathbf{x} + Res, \tag{3.17}$$

where the quantity Res does not depend on $\mathbf{p}$, $\beta = -\lambda_3^k \cdot \mathbf{n}^k - \lambda(|\mathbf{m}^k| - |\nabla R|)$ and

$$C = \frac{r_2\nabla K^k + r_3|\mathbf{p}^k|\mathbf{n}^k + r_4(\mathbf{m}^k - \nabla R) - \lambda_2^k - \lambda_3^k - \lambda_4^k}{r_2 + r_3 + r_4 + \lambda}. \tag{3.18}$$

The minimization problem (3.17) has a closed from solution which is explicitly given by the following shrinkage-like formula:

$$\mathbf{p}^{k+1} = \max\left\{1 - \frac{\beta}{(r_2 + r_3 + r_4 + \lambda)|C|}, 0\right\} C. \tag{3.19}$$

The n-subproblem. Fixing $\mathbf{u}^{k+1}$, K^{k+1} and $\mathbf{p}^{k+1}$ and λ_i^k $(i = 1, \cdots, 5)$, the $\mathbf{n}$-problem consists in solving the following minimization problem:

$$\min_{\mathbf{n}} \frac{\lambda}{2}\int_\Omega (\mathbf{n} - \nabla_n R)^2 d\mathbf{x} + \frac{r_3}{2}\int_\Omega (\mathbf{p}^{k+1} - |\mathbf{p}^{k+1}|\mathbf{n})^2 d\mathbf{x} + \int_\Omega (\mathbf{p}^{k+1} - |\mathbf{p}^{k+1}|\mathbf{n}) \cdot \lambda_3^k d\mathbf{x}.$$

The above problem has a closed from solution which is is explicitly given by:

$$\mathbf{n} = \frac{\lambda\nabla_n R + r_3|p^{k+1}|p^{k+1} + |\mathbf{p}^{k+1}|\lambda_3^k}{\lambda + r_3}. \tag{3.20}$$

The m-subproblem. To find the optimal value of $\mathbf{m}^{k+1}$, we solve the following optimisation sub-problem:

$$\begin{aligned}\min_{\mathbf{m}} \frac{\lambda}{2}\int_\Omega (|\mathbf{p}^{k+1}| + |\nabla R| - |\mathbf{m}|)^2\, d\mathbf{x} + \frac{r_4}{2}\int_\Omega (\mathbf{p}^{k+1} + \nabla R - \mathbf{m})^2\, d\mathbf{x} \\ + \int_\Omega (\mathbf{p}^{k+1} + \nabla R - \mathbf{m})\cdot \lambda_4^k\, d\mathbf{x}.\end{aligned} \tag{3.21}$$

The above problem is equivalent to minimizing the following energy:

$$\min_{\mathbf{m}} -\lambda\int_\Omega (|\mathbf{p}^{k+1}| + |\nabla R|)|\mathbf{m}|\, d\mathbf{x} + \frac{\lambda + r_4}{2}\int_\Omega (\mathbf{m} - C)^2 d\mathbf{x} + Res,$$

where both Res and C do not depend on $\mathbf{p}$, with C given by:

$$C = \frac{r_4(\mathbf{p}^{k+1} + \nabla R) + \lambda_4^k}{\lambda + r_4}.$$

The solution is explicitly given by:

$$\mathbf{m}^{k+1} = \max\left\{1 + \frac{\lambda(|\mathbf{p}^{k+1}| + |\nabla R|)}{(\lambda + r_4)|C|}, 0\right\} C. \tag{3.22}$$

Lemma 3.1 ([37]). *Let $f : \mathbb{R} \to \mathbb{R}$ be a closed, proper and convex function. Let $(w_n)_{n\in\mathbb{N}}$ be a sequence of distinct functions in dom f converging to $w^* \in int(domf)$ and let $S_n \in \partial f(w_n)$. Then there exists a subsequence $(S_{n_k})_{k\in\mathbb{N}}$ that converges to the point S^*, where $S^* \in \partial f(w^*)$.*

In the sequel, we give a partial result about the limit behaviour of the solutions generated by the ALM method. Let us consider the space:

$$\mathcal{X} = \tilde{\mathcal{W}} \times W_0^{1,2}(\Omega) \times L^2_{\text{div}}(\Omega) \times L^2_{\text{div}}(\Omega) \times L^2(\Omega) \times L^2_{\text{div}}(\Omega) \times L^2_{\text{div}}(\Omega) \times L^2_{\text{div}}(\Omega) \times L^2(\Omega),$$

where $\tilde{\mathcal{W}} = \{u \in \mathcal{W}, \text{div}^2.\nabla^2 u \in L^2(\Omega)\}$ and

$$L^2_{\text{div}}(\Omega) = \{w \in (L^2(\Omega))^2,\, \text{div}\, w \in L^2(\Omega)\}.$$

Proposition 3.2. *If the sequence $(\mathbf{u}^k, K^k, \mathbf{p}^k, \mathbf{n}^k, \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k, \lambda_5^k) \in \mathcal{X}$, generated by the ALM method, converges to a point $(\mathbf{u}^*, K^*, \mathbf{p}^*, \mathbf{n}^k, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*) \in \mathcal{X}$, then the limit point satisfies the following first-order optimality conditions:*

$$\begin{cases} -\alpha\Delta u_1^* + \alpha_1 div^2.\nabla^2 u_1^* + \lambda_1\partial_x T(\mathbf{u}^*) + \partial_{u_1}\mathcal{C}_s(\mathbf{u}^*, \lambda_5^*) = 0, & in\ \Omega, \\ -\alpha\Delta u_2^* + \alpha_1 div^2.\nabla^2 u_2^* + \lambda_1\partial_x T(\mathbf{u}^*) + \partial_{u_2}\mathcal{C}_s(\mathbf{u}^*, \lambda_5^*) = 0, & in\ \Omega, \\ div\,\lambda_2^* - \lambda_1^* = 0, \qquad -\beta^* S_{\mathbf{p}}^* + \sum_{i=2}^4 \lambda_i^* = 0, \\ \lambda\mathbf{n}^* - \lambda\nabla_n R - |\mathbf{p}^*|\lambda_3^* = 0, \qquad -\lambda(|\mathbf{p}^*| + |\nabla R|)S_{\mathbf{m}}^* + \lambda\mathbf{m}^* - \lambda_4^* = 0 \\ \min(\lambda_5^*, \sigma\mathcal{F}(\mathbf{u}^*)) = 0, \qquad T(\mathbf{u}^*) = K^*, \quad \mathbf{p}^* = \nabla K^*, \\ \mathbf{m}^* = \mathbf{p}^* + \nabla R, \qquad \mathbf{p}^* = |\mathbf{p}^*|\mathbf{n}^*, \end{cases}$$

where $\beta^* = -\lambda_3^* \cdot \mathbf{n}^* - \lambda(|\mathbf{m}^*| - |\nabla R|)$. Consequently $\mathbf{u}^* = (u_1^*, u_2^*)$ is a stationary point of model (2.1).

Proof. By (3.5), (3.6), (3.7) and (3.8), we have:

$$\lim \frac{1}{r_1}(\lambda_1^{k+1} - \lambda_1^k) = \lim(T(\mathbf{u}^{k+1}) - K^{k+1}) = T(\mathbf{u}^*) - K^* = 0, \tag{3.23}$$

$$\lim \frac{1}{r_2}(\lambda_2^{k+1} - \lambda_2^k) = \lim(\mathbf{p}^{k+1} - \nabla K^{k+1}) = \mathbf{p}^* - \nabla K^* = 0, \tag{3.24}$$

$$\lim \frac{1}{r_3}(\lambda_3^{k+1} - \lambda_4^k) = \lim(\mathbf{n}^{k+1} - |\mathbf{p}^{k+1} n^{k+1}) = \mathbf{n}^* - |\mathbf{p}^*| n^* = 0. \tag{3.25}$$

$$\lim \frac{1}{r_4}(\lambda_4^{k+1} - \lambda_4^k) = \lim(\mathbf{m}^{k+1} - |\mathbf{p}^{k+1}| - \nabla R) = \mathbf{m}^* - \mathbf{p}^* - \nabla R = 0. \tag{3.26}$$

From (3.8), we get:

$$0 = \lim(\lambda_5^{k+1} - \lambda_5^k) = \lim \max\{-\lambda_5^k, -\sigma \mathcal{F}_\epsilon(\mathbf{u}^{k+1})\} = \min\{\lambda_5^*, \sigma \mathcal{F}_\epsilon(\mathbf{u}^*)\}. \tag{3.27}$$

Back to the optimality condition for the $\mathbf{u}$-subproblem in (3.11), taking the limit in (3.11) over k and considering equalities (3.24) and (3.26), we get:

$$\begin{cases} -\alpha \Delta u_1^* + \alpha_1 \mathrm{div}^2 . \nabla^2 u_1^* + \lambda_1 \partial_x T(\mathbf{u}^*) + \partial_{u_1} \mathcal{C}_s(\mathbf{u}^*, \lambda_5^*) = 0, & in \;\; \Omega, \\ -\alpha \Delta u_2^* + \alpha_1 \mathrm{div}^2 . \nabla^2 u_2^* + \lambda_1 \partial_x T(\mathbf{u}^*) + \partial_{u_2} \mathcal{C}_s(\mathbf{u}^*, \lambda_5^*) = 0, & in \;\; \Omega. \end{cases}$$

Now, we consider the optimality conditions for the K-subproblem and take the limit over k:

$$\begin{aligned} &- r_2(\Delta K^{k+1} - \mathrm{div}\mathbf{p}^k) + r_1(K^{k+1} - T(\mathbf{u}^{k+1})) + \mathrm{div}\, \lambda_2^k - \lambda_1^k = 0, \text{ i.e.} \\ &- r_2 \mathrm{div}(\nabla K^* - \mathbf{p}^*) + r_1(K^* - T(\mathbf{u}^*)) + \mathrm{div}\, \lambda_2^* - \lambda_1^* = 0 \end{aligned}$$

where we used $\mathrm{div}\nabla K^* = \Delta K^*$. Using the equalities (3.23) and (3.24), $\nabla K^* - \mathbf{p}^* = 0$ and $K^* - T(\mathbf{u}^*) = 0$. Then $\mathrm{div}\lambda_2^* - \lambda_1^* = 0$. The optimality condition for the modified $\mathbf{p}$-subproblem (3.17) leads to:

$$-\beta S_{\mathbf{p}}^{k+1} + r_2(\mathbf{p}^{k+1} - \nabla K^{k+1}) + r_3(\mathbf{p}^{k+1} - |\mathbf{p}^k| n^k)) + r_4(\mathbf{p} + \nabla R - \mathbf{m}^k) + \sum_{i=2}^{4} \lambda_i^k,$$

where $S_{\mathbf{p}}^{k+1} \in \partial |\mathbf{p}^{k+1}|$ and β is given in (3.18). By Lemma 3.1, there exists a subsequence, still denoted by $S_{\mathbf{p}}^k \in \partial |\mathbf{p}^k|$, converging to $S_{\mathbf{p}}^* \in \partial |\mathbf{p}^*|$. Taking the limit over k and taking into account equalities (3.24) and (3.25), we obtain:

$$-\beta^* S_{\mathbf{p}}^* + \sum_{i=2}^{4} \lambda_i^* = 0.$$

For the $\mathbf{n}$-subproblem, the optimality conditions give:

$$\lambda(\mathbf{n}^{k+1} - \nabla_n R) + r_3(\mathbf{p}^{k+1} - |\mathbf{p}^{k+1}| \mathbf{n}^{k+1})^2 + \lambda_3^k d\mathbf{x} = 0$$

Considering the limit over k (3.25), we get:

$$\lambda \mathbf{n}^* - \lambda \nabla_n R - |\mathbf{p}^*| \lambda_3^* = 0.$$

The same analysis applied to the optimality condition for the $\mathbf{m}$-subproblem (3.21) leads to the equality:

$$-\lambda(|\mathbf{p}^*| + |\nabla R|) S_{\mathbf{m}}^* + \lambda \mathbf{m}^* - \lambda_4^* = 0, \; S_{\mathbf{m}}^* \in \partial |\mathbf{m}^*|.$$

∎

Finally we remark on getting the initializations by a multiresolution technique, also to avoid local minima and to speed up registration. We use a scale space approach by resizing the original images to a sequence of coarser ones where computations are cheap and register these smaller images (see Fig. 3). Then starting from the coarsest level, we interpolate the obtained transformation fields to get a starting guess on finer (next) levels until the original resolution on the finest level is reached.

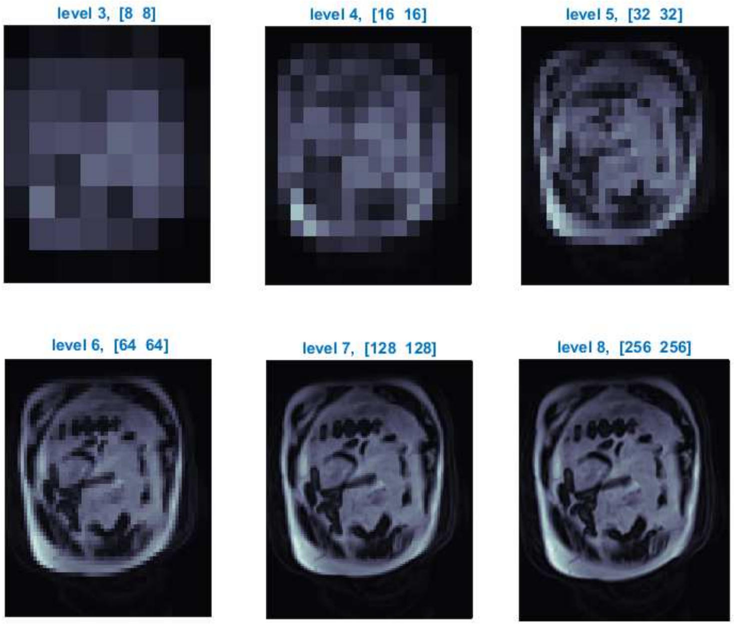


Figure 3. *Example of a multilevel representation of images.*

4. Numerical experiments.

In this section, we assess the performance of the proposed model (denoted by "**New Model**" below) and its algorithm. We compare the proposed model with two other multimodality models:

- A MI model (denoted by **MI** below) that combines the regulariser (2.2) and the MI similarity measure (1.3);
- A NGF model (denoted by **NGF** below) that combines the regulariser (2.2) and the standard NGF similarity measure (1.4).

To measure the quality of the registered images, the following quantity

$$\mathbf{GF_{er}} = \frac{F(\nabla T(\mathbf{u}), \nabla R)}{F_0} \tag{4.1}$$

is used as the relative reduction of the dissimilarity, where for two vectors $x = (x_1, x_2)$ and $y = (y_1, y_2)$, we have

$$F(x, y) = \| \frac{x_t}{\|x_t\|} - \frac{y_t}{\|y_t\|} \|_1, \; x_t = (x_1, x_2), \;\; y_t = (y_1, y_2).$$

Here, $F_0 = F(\nabla T(\mathbf{u}), \nabla R)$ if $\mathbf{u} = \mathbf{0}$. For additional criteria to measure the goodness of registration, we also use the relative normalized gradient fields

$$\mathbf{NGF_{er}} = \frac{D^{NGF}(T(\mathbf{u}), R)}{NGF_0} \tag{4.2}$$

where $NGF_0 = D^{NGF}(T(\mathbf{u}), R)$ if $\mathbf{u} = \mathbf{0}$, and non-negative mutual information measure

$$\mathbf{MI_{er}} = -D^{MI}(T(\mathbf{u}), R). \tag{4.3}$$

For all the numerical experiments presented here, we summarise the comparative results in a table where we give the error computed using formulas (4.1)-(4.2). To measure mesh validity, we compute $\mathcal{C}(\mathbf{u}) = \det\,(I + \nabla \mathbf{u})$ from (2.1) and monitor if it is positive.

In order to reduce the number of parameters to tune, we set $r_1 = 5$, $r_2 = 10$ and $r_3 = r_4 = 100$ in all numerical experiments unless stated otherwise. We consider $N_{max} = 70$ the maximum number of iterations for **New Model** from Algorithm 3.2 and we stop the iterations before reaching $N_{max} = 70$ if the following stopping criterion

$$\frac{\|\mathbf{p}^k + \nabla R - \mathbf{m}^k\|_{L^1}}{\sqrt{l \times c}} \leq \tau$$

is satisfied for a given tolerance $\tau = 10^{-3}$, where l and c are the numbers of rows and columns in the image. Though we can use all equations from Algorithm 3.1 to stop iterations, we find that the above stopping criterion based on its 4th equation is sufficient as it includes information about the gradients of both images. Thus, it can control the ALM iterations and the quality of registration at the same time. For each variable u_1 and u_2, we computed the residual via finite differences approximation and the global residual is taken as the sum. The residual is given by the quantity

$$S_{er} = \frac{1}{l \times c} \sum_{i=1}^{l} \sum_{j=1}^{c} \Big| \Big(\frac{\partial \mathcal{J}_1(\mathbf{u}^k)}{\partial u_1^k} \Big)_{i,j} \Big| + \frac{1}{l \times c} \sum_{i=1}^{l} \sum_{j=1}^{c} \Big| \Big(\frac{\partial \mathcal{J}_1(\mathbf{u}^k)}{\partial u_2^k} \Big)_{i,j} \Big| \tag{4.4}$$

where $(\dfrac{\partial \mathcal{J}_1(\mathbf{u}^k)}{\partial u_1^k})_{i,j} = \dfrac{J(\mathbf{u}^k) - J(\mathbf{u}^{kij})}{u_1^k(i,j) - u_1^{k-1}(i,j)}$, $\mathbf{u}^{kij} = (u_1^{kij}, u_2^k)$ and u_1^{kij} takes the same values of u_1^k on each point of the discrete domain, except on the position (i,j) where it takes the values of the old ALM solution $u_1^{k-1}(i,j)$. The term $(\dfrac{\partial \mathcal{J}_1(\mathbf{u}^k)}{\partial u_2^k})_{i,j}$ is defined in a similar way. We also plot the curve of the quantity

$$D_m = \frac{D^{GF}(T(\mathbf{u}), R)}{D^{GF}(T, R)} + \frac{D^{TM}(T(\mathbf{u}), R)}{D^{TM}(T, R)} \tag{4.5}$$

which represents the relative errors for the new similarity measures as function of the ALM iterations.

For the **NGF** and **MI** similarity measures, the numerical experiments are performed using the publicly available image registration toolbox flexible algorithms for image registration (FAIR)[1], where the implementation is based on the Gauss-Newton method. The constraint

[1] http://www.siam.org/books/fa06/

on the determinant $\det(I + \nabla \mathbf{u}) > 0$ is explicitly included in FAIR's models; in fact, a line search method is used in FAIR and the new descent direction is chosen such that the constraint $\det(I + \nabla \mathbf{u}) > 0$ is verified.

As we shall see, in almost all experiments, the **New Model** outperforms the standard **NGF** and the **New Model** also outperforms **MI** in examples where dominating gradients represent main image features or they correspond to each other, while the **New Model** performs similarly to **MI** for other examples (e.g. Example 6).

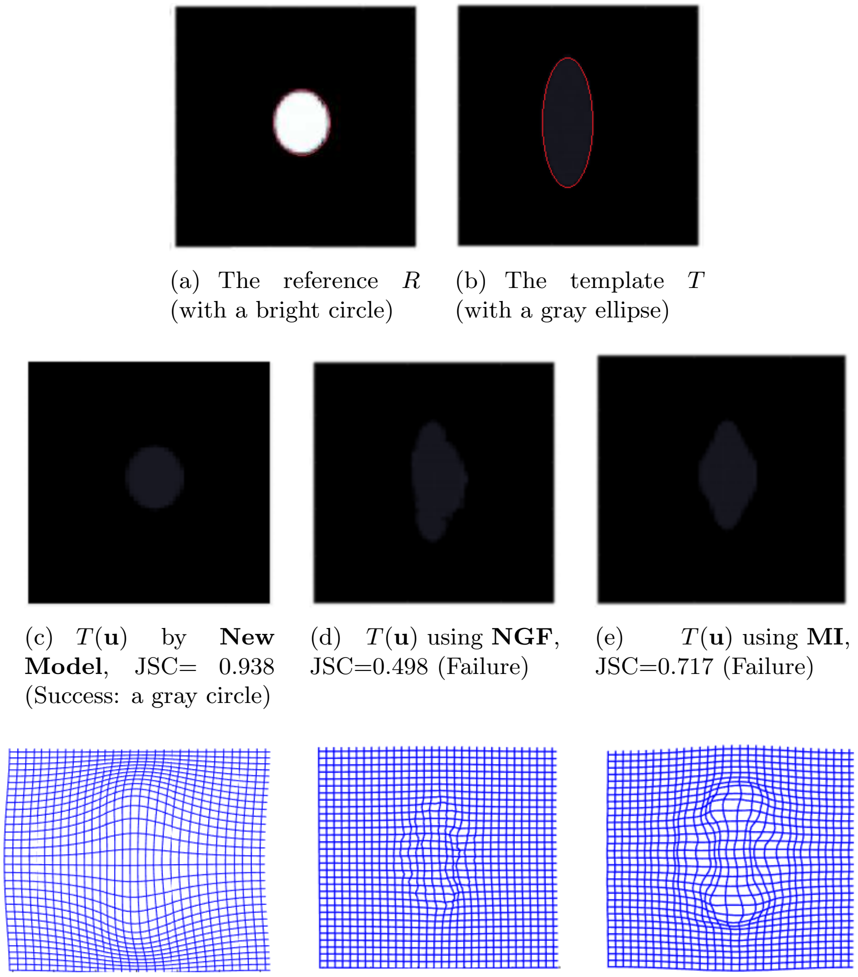
Example 1. In the first example, we consider a synthetic image to illustrate the type of images where mutual information (**MI**) and the normalized gradient field (**NGF**) models are at disadvantages. We obtain a good result using **New Model** as seen in Fig.4. Here, the **NGF** and **MI** models were tested for different regularization parameters. The optimal choices are considered by making different tests where we set $\alpha_1 = 1$, $\alpha = 0.01\alpha_1$ and we vary λ such that $\frac{\alpha_1}{\lambda} \in \{10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 10^{-2}\}$ for **MI**, and $\frac{\alpha_1}{\lambda} \in \{2.5, 2, 1.5, 1, 0.5, 0.1\}$ for **NGF**. The optimal parameters were $\frac{\alpha_1}{\lambda} = 10^{-4}$ and $\frac{\alpha_1}{\lambda} = 10^{-4} = 0.5$ for **MI** and **NGF**, respectively. They were chosen such that the registered image is very close to the reference and the transformations does not suffer from mesh folding. For comparison, we used the Jaccard similarity coefficient (JSC) which is defined as follows:

$$JSC = \frac{|S_T \cap S_R|}{|S_T \cup S_R|}, \tag{4.6}$$

where S_T and S_R represent, respectively, the segmented regions of interest (with red contour) in the deformed template (after registration) and the reference.

Examples 2 and 3. In Fig 5, we consider a reference image from photon density weighted MRI and a template image which represent MRI-T2, both of size 256×256. A seocnd set of examples is shown in Fig 6. We compare with the different multi-modal registration models. For each model, we display registered templates. We can see that all models perform well for both examples and give satisfactory results. The results of the **NGF** and **MI** are broadly comparable. In both examples with all models, the results for the registration look visually identical. We display an overlay in alternating squared patches of the registered and the reference image (to possibly see major discontinuities of features). We quantify the quality of registration using the $\mathbf{GF_{er}}$, $\mathbf{MI_{er}}$ and $\mathbf{NGF_{er}}$ errors which confirm that **New Model**; e.g. at the top left (second box down) of Fig 5, gives better alignments than compared models. For the run runtime comparison with the **MI** and **NGF** models, we tested all models for the pair images in 5 for different resolutions. The FAIR's models are always slightly faster because they are optimized (based on Gauss-Newton method)

Example 4. In Fig. 8, we present the result of registering two diffusion-MRI images of size 256×256 with respectively high and low b-value diffusion. Since the intensity values for different b-values are not comparable, conventional non-modality registration models (that rely on matching the images based on the intensity values) will fail. We show the registration results by our compared 3 models in Fig. 8. We notice that **NGF** and **MI** models give comparable results. However, our **New Model** gives the best result comparing to the other two and visually, the reference and the transformed template are well aligned in all regions. Since $\mathcal{C}(\mathbf{u}) > 0$, all transformed grids have no mesh folding.



(a) The reference R (with a bright circle)

(b) The template T (with a gray ellipse)

(c) $T(\mathbf{u})$ by **New Model**, JSC= 0.938 (Success: a gray circle)

(d) $T(\mathbf{u})$ using **NGF**, JSC=0.498 (Failure)

(e) $T(\mathbf{u})$ using **MI**, JSC=0.717 (Failure)

(f) $x + \mathbf{u}(x)$ using **New Model**

(g) $x+\mathbf{u}(x)$ using **NGF**

(h) $x + \mathbf{u}(x)$ using **MI**

Figure 4. *Example 1: Comparison of three different models. Clearly only* **Our Model** *works while* **NGF**, **MI** *fail completely.*

Example 5. In the next experiment in Fig. 9, our aim is to investigate capabilities of the proposed models for registration of MRI-T1 and MRI-T2 images in higher resolution 512×512. We can observe from overlaying of the registered and the reference images that all models work fine in producing acceptable registration results, however the registered result by **New Model** produces the best alignment in all parts and gives the better similarity value than **NGF** (here identical to **MI**). We also show the resulting transformed grids for all models where there is no mesh folding due to $\mathcal{C}(\mathbf{u}) > 0$. For the above 4 examples (Ex.2–Ex.5), in Fig 10, we display the evolution of the error versus the ALM iteration to the final solution. We also plot the evolution of the residual for the energy (2.1) as a function of ALM iterations. Here we see that our ALM algorithm converges though the convergence is not monotone.

Example 6. Example 6 tests the registration of a MRI image to a PET with much noise In Fig. 12, we present the results obtained using the **New Model**, **NGF** and **MI**. Clearly, **New Model** and **MI** perform better than **NGF** in this case and in particular the **New Model** performs the best (even though it is slightly better than **MI Model**). We display an overlaying of the registered and the reference images which shows that the registered result by **New Model** produces the best alignment.

Regularisation parameters dependence test. In Table 3, we compare the sensitivity of the proposed model with respect to varying the ratio $\frac{\alpha_1}{\lambda}$. The model was tested on Example 2 where we set $\alpha_1 = 1$, $\alpha = 0.01\alpha_1$ and we vary λ for all experiments. We can see a clear process of the changes of the relative error where the best error is obtained for $\frac{\alpha_1}{\lambda} = 0.017$ and the error increases as the ratio decreases more than 0.017.

	Resolution			
	64×64	128×128	256×256	512×512
Time (s) for **New Model**	29.836	49.931	117.342	272.578
Time (s) for **MI Model**	14.794	21.437	48.881	76.398
Time (s) for **NGF Model**	22.003	42.845	100.961	264.388

Table 1

Run time comparison for all models for the pair of MRI images in Fig. 6

5. Conclusions.

Image registration is an increasingly important and often challenging image processing task with a broad range of applications such as in astronomy, optics, biology, chemistry and medical imaging. In this paper to improve the multi-modality registration model based on the normalized gradients of the images, we propose a new gradients-based variational model using a regularisation term which combines first- and second-order derivatives of the displacement. After showing the solution existence, we present a fast ALM for its numerical implementation. Experimental tests confirm that our proposed model performs better in multi-modality images registration than compared models. It is pleasing to see much improved results over established models within the same modelling framework. Future work will consider generalizations to 3 dimensions and registration of images that do not have dominant gradients.

REFERENCES

[1] E. Bae, J. Shi, and X.-C. Tai, *Graph cuts for curvature based image denoising*, IEEE Transactions on Image Processing, 20 (2011), pp. 1199–1210.

[2] M. Burger, J. Modersitzki, and L. Ruthotto, *A hyperelastic regularization energy for image registration*, SIAM Journal on Scientific Computing, 35 (2013), pp. 132–148.

[3] Y. M. Chen, J. L. Shi, M. Rao, and J. S. Lee, *Deformable multi-modal image registration by maximizing renyi's statistical dependence measure*, Inverse Problems and Imaging, 9 (2015), pp. 79–103.

[4] N. Chumchob, *Vectorial total variation-based regularization for variational image registration*, IEEE Transactions on Image Processing, 22 (2013), pp. 4551–4559.

[5] N. Chumchob and K. Chen, *Improved variational image registration model and a fast algorithm for its numerical approximation*, Numerical Methods for Partial Differential Equations, 28 (2012), pp. 1966–1995.

Compared Models			NGF			New Model			MI		
	#G	#N	$\mathbf{GF_{er}}$	$\mathbf{NGF_{er}}$	$\mathbf{MI_{er}}$	$\mathbf{GF_{er}}$	$\mathbf{NGF_{er}}$	$\mathbf{MI_{er}}$	$\mathbf{GF_{er}}$	$\mathbf{NGF_{er}}$	$\mathbf{MI_{er}}$
Ex 1	0.2%	.02%	0.540	0.964	0.446	0.032	0.932	0.993	0.370	0.97	0.381
Ex 2	49%	24%	0.636	0.640	1.170	0.247	0.756	1.206	0.490	0.879	1.193
Ex 3	49%	23%	0.336	0.491	1.265	0.238	0.389	1.290	0.463	0.579	1.265
Ex 4	49%	20%	0.901	0.856	1.150	0.674	0.800	1.184	0.765	0.849	1.154
Ex 5	43%	37%	0.741	0.656	1.163	0.454	0.623	1.178	0.454	0.631	1.163
Ex 6	48%	23%	0.952	0.957	1.187	0.801	0.920	1.341	0.836	0.970	1.254

Table 2

Registration results of the different models for processing Examples 1-5 shown respectively in Fig. 5, 6, 7 and 8. The errors are computed using formula (4.1), (4.3) *and* (4.2). *Here, #N is the ratio of the number of pixels where* $\nabla_n T \cdot \nabla_n R \neq 0$ *over the total number of pixels, whereas #G is the ratio of number pixels where* $\mathbf{GF}(T, R) + \mathbf{TM}(T, R) \neq 0$ *over the total number of pixels.*

$\frac{\alpha_1}{\lambda}$	0.1	0.05	0.025	0.017	0.0125	0.01	0.0075	0.005
Error	0.238	0.237	0.237	0.236	0.237	0.237	0.238	0.24

Table 3

Registration results for $\frac{\alpha_1}{\lambda}$*-dependence tests of* **New Model** *for processing Example 3. The relative errors are computed using the normalized gradient fitting formula* (4.1). *In all cases, we set* $\alpha = 0.01\alpha_1$.

[6] N. Chumchob, K. Chen, and C. Brito-Loeza, *A fourth-order variational image registration model and its fast multigrid algorithm*, Multiscale Modeling & Simulation, 9 (2011), pp. 89–128.

[7] M. Droske and W. Ring, *A mumford–shah level-set approach for geometric image registration*, SIAM journal on Applied Mathematics, 66 (2006), pp. 2127–2148.

[8] J. Feydy, B. Charlier, F. V. Vialard, and G. Peyre, *Optimal transport for diffeomorphic registration*, https://arxiv.org/abs/1706.05218v1, (2017).

[9] B. Fischer and J. Modersitzki, *Fast diffusion registration*, Contemp. Math., 313 (2002), pp. 117–129.

[10] B. Fischer and J. Modersitzki, *Curvature based image registration*, Journal of Mathematical Imaging and Vision, 18 (2003), pp. 81–85.

[11] ———, *Ill-posed medicine - an introduction to image registration*, Inverse Problems, 24 (2008).

[12] E. Haber and J. Modersitzki, *Numerical methods for volume preserving image registration*, Inverse problems, 20 (2004), pp. 1621–1638.

[13] ———, *Image registration with guaranteed displacement regularity*, International Journal of Computer Vision, 71 (2007), pp. 361–372.

[14] S. Henn, *A multigrid method for a fourth-order diffusion equation with application to image processing*, SIAM Journal on Scientific Computing, 27 (2005), pp. 831–849.

[15] E. Hodneland, A. Lundervold, J. Rørvik, and A. Z. Munthe-Kaas, *Normalized gradient fields for nonlinear motion correction of dce-mri time series*, Computerized Medical Imaging and Graphics, 38 (2014), pp. 202–210.

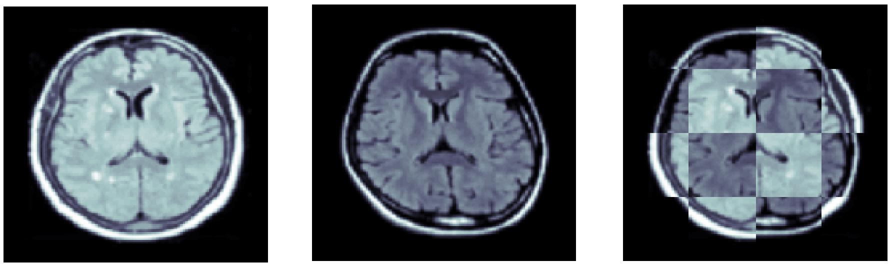
[16] W. Hu, Y. Xie, L. Li, and W. Zhang, *A total variation based nonrigid image registration by combining parametric and non-parametric transformation models*, Neurocomputing, 144 (2014), pp. 222 – 237.

[17] M. Ibrahim, K. Chen, and C. Brito-Loeza, *A novel variational model for image registration using gaussian curvature*, Geometry, Imaging and Computing, 1 (2014), pp. 417 – 446.

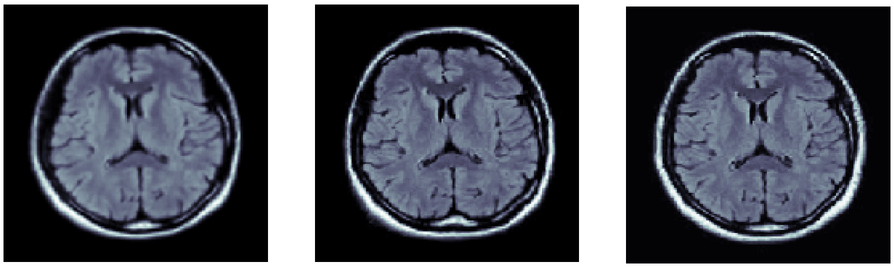
[18] L. König and J. Rühaak, *A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields*, in Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, IEEE, 2014, pp. 580–583.

[19] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, *Nonrigid image registration using conditional mutual information*, IEEE transactions on medical imaging, 29 (2010), pp. 19–29.
[20] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, *Multimodality image registration by maximization of mutual information*, IEEE Transactions on Tedical Imaging, 16 (1997), pp. 187–198.
[21] A. Mang and G. Biros, *An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 1030–1069.
[22] ———, *Constrained h^1-regularization schemes for diffeomorphic image registration*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1154–1194.
[23] J. Modersitzki, *FAIR: Flexible Algorithms for Image Registration*, SIAM, 2009.
[24] F. P. Oliveira and J. M. R. Tavares, *Medical image registration: a review*, Computer methods in biomechanics and biomedical engineering, 17 (2014), pp. 73–93.
[25] K. Papafitsoros, C. B. Schoenlieb, and B. Sengul, *Combined first and second order total variation inpainting using split bregman*, Image Processing On Line, 3 (2013), pp. 112–136.
[26] J. P. Pluim, J. A. Maintz, and M. A. Viergever, *Mutual-information-based registration of medical images: a survey*, IEEE transactions on medical imaging, 22 (2003), pp. 986–1004.
[27] C. Pöschl, J. Modersitzki, and O. Scherzer, *A variational setting for volume constrained image registration*, Inverse Problems and Imaging, 4 (2010), pp. 505–522.
[28] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs, *Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint*, IEEE transactions on medical imaging, 22 (2003), pp. 730–741.
[29] G. Roland and L. T. Patrick, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, SIAM, 1989.
[30] J. Rühaak, L. König, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, and B. Fischer, *A fully parallel algorithm for multimodal image registration using normalized gradient fields*, in Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, IEEE, 2013, pp. 572–575.
[31] A. Sotiras, C. Davatzikos, and N. Paragios, *Deformable medical image registration: A survey*, IEEE Transactions on Medical Imaging, 32 (2013), pp. 1153–1190.
[32] X.-C. Tai, J. Hahn, and G. J. Chung, *A fast algorithm for Euler's elastica model using augmented lagrangian method*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 313–344.
[33] P. Viola and W. M. Wells III, *Alignment by maximization of mutual information*, International Journal of Computer Vision, 24 (1997), pp. 137–154.
[34] C. Wu and X. C. Tai, *Augmented lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 300–339.
[35] C. Wu, J. Zhang, and X.-C. Tai, *Augmented lagrangian method for total variation restoration with non-quadratic fidelity*, Inverse problems and imaging, 5 (2011), pp. 237–261.
[36] C. Xing and P. Qiu, *Intensity-based image registration by nonparametric local smoothing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2011), pp. 2081–2092.
[37] M. Yashtini and S. H. Kang, *A fast relaxed normal two split method and an effective weighted TV approach for E1uler's elastica image inpainting*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1552–1581.
[38] W. Yilun, Y. Junfeng, Y. Wotao, and Z. Yin, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 248–272.
[39] J. Zhang and K. Chen, *Variational image registration by a total fractional-order variation model*, Journal of Computational Physics, 293 (2015), pp. 442–461.
[40] J. Zhang, K. Chen, and B. Yu, *An improved discontinuity-preserving image registration model and its fast algorithm*, Applied Mathematical Modelling, 40 (2016), pp. 10740–10759.
[41] ———, *A novel high-order functional based image registration model with inequality constraint*, Computers & Mathematics with Applications, 72 (2016), pp. 2887–2899.
[42] X. Zhou, *Weak lower semicontinuity of a functional with any order*, Journal of Mathematical Analysis and Applications, 221 (1998), pp. 217–237.
[43] W. ZHU, X.-C. TAI, and T. CHAN, *Augmented lagrangian method for a mean curvature based image denoising model.*, Inverse Problems & Imaging, 7 (2013).

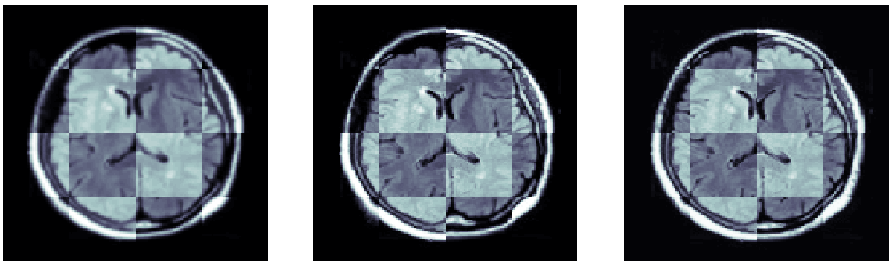
[44] W. Zhu, X.-C. Tai, and T. Chan, *Image segmentation using eulers elastica as the regularization*, Journal of Scientific Computing, 57 (2013), pp. 414–438.



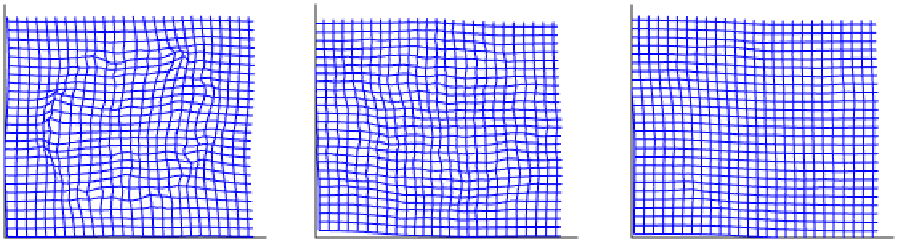
(a) The reference R (b) The template T (c) Overlay of R and T



(d) $T(\mathbf{u})$ by **New Model**, $\mathbf{GF_{er}}$=0.247,$\mathbf{MI_{er}}$=1.206, $\mathbf{NGF_{er}}$= 0.756 (e) $T(\mathbf{u})$ using **NGF**, $\mathbf{GF_{er}}$=0.636, $\mathbf{MI_{er}}$=1.17, $\mathbf{NGF_{er}}$=0.640 (f) $T(\mathbf{u})$ using **MI**, $\mathbf{GF_{er}}$=0.490, $\mathbf{MI_{er}}$=1.193, $\mathbf{NGF_{er}}$=0.879

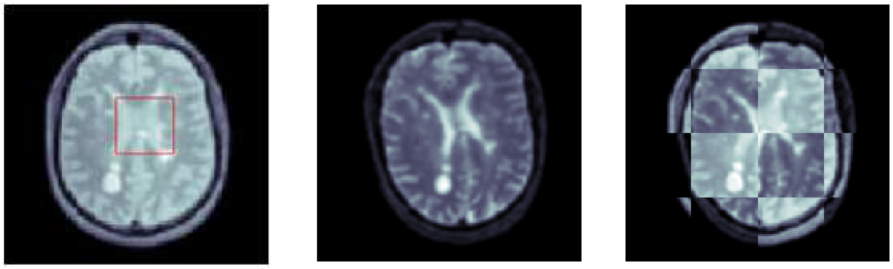


(g) Overlay of $R, T(\mathbf{u})$ for **New Model** (h) Overlay of R and $T(\mathbf{u})$ for **NGF** (i) Overlay of R and $T(\mathbf{u})$ for **MI**



(j) $x + \mathbf{u}(x)$ by **New Model**, $\min \mathcal{C} = 0.22$ (k) $x + \mathbf{u}(x)$ using **NGF**, $\min \mathcal{C} = 0.19$ (l) $x + \mathbf{u}(x)$ using **MI**, $\min \mathcal{C} = 0.45$

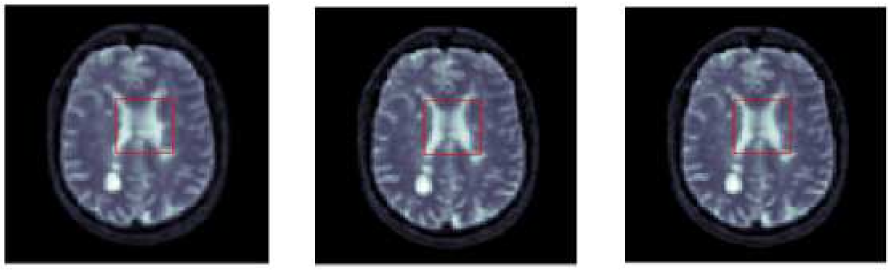
Figure 5. *Example 2: Comparison of different models to register T-1 and T2-MRI images.* **New Model** *performs the best.*



(a) The reference R

(b) The template T

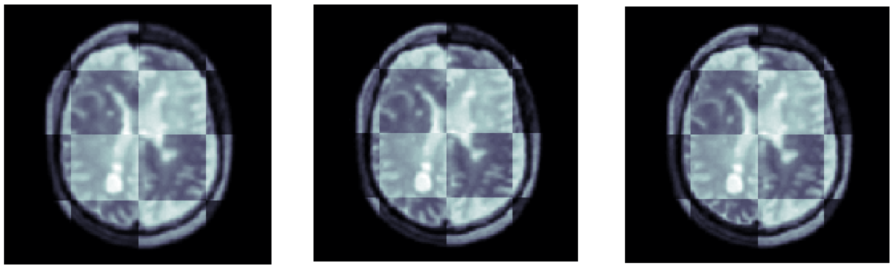
(c) Overlay of R and T



(d) $T(\mathbf{u})$ by **New Model**, $\mathbf{GF_{er}}$=0.278, $\mathbf{MI_{er}}$=1.290, $\mathbf{NGF_{er}}$=0.389

(e) $T(\mathbf{u})$ using **NGF**, $\mathbf{GF_{er}}$=0.336, $\mathbf{MI_{er}}$=1.265, $\mathbf{NGF_{er}}$=0.491

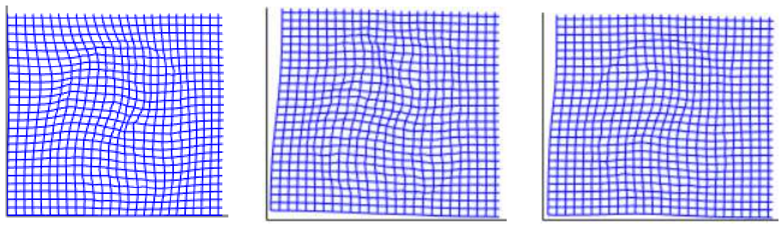
(f) $T(\mathbf{u})$ using **MI**, $\mathbf{GF_{er}}$=0.463, $\mathbf{MI_{er}}$=1.265, $\mathbf{NGF_{er}}$= 0.579



(g) Overlay of $R, T(\mathbf{u})$ for **New Model**

(h) Overlay of R and $T(\mathbf{u})$ for **NGF**

(i) Overlay of R and $T(\mathbf{u})$ for **MI**



(j) $x + \mathbf{u}(x)$ by **New Model**, $\min \mathcal{C} = 0.31$

(k) $x + \mathbf{u}(x)$ using **NGF**, $\min \mathcal{C} = 0.31$

(l) $x + \mathbf{u}(x)$ using **MI**, $\min \mathcal{C} = 0.33$

Figure 6. *Example 3: Registration of a second pair of MRI images (T1 and T2).* **New Model** *performs the best.*

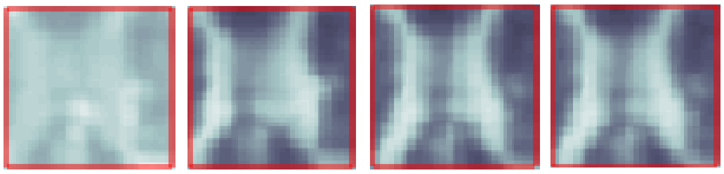
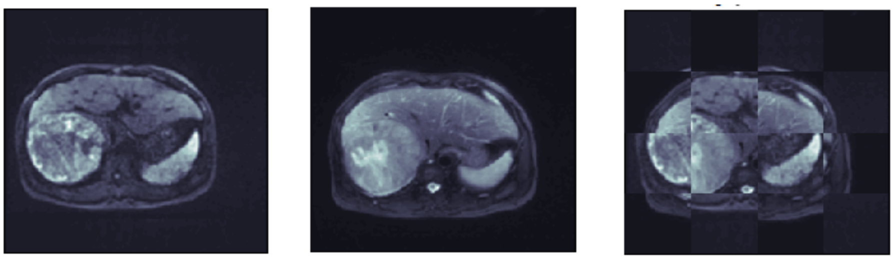
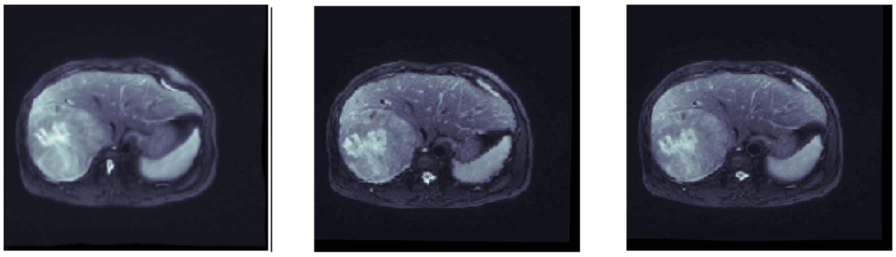


Figure 7. *Comparison of* 3 *different models to register the MRI images fin Fig. 6. Example 3 zoomed in the red squares (see Fig. 6): From left to right; Zooms in the reference R and the registered $T(\mathbf{u})$ using* **New model**, **NGF** *and* **MI**, *respectively.*



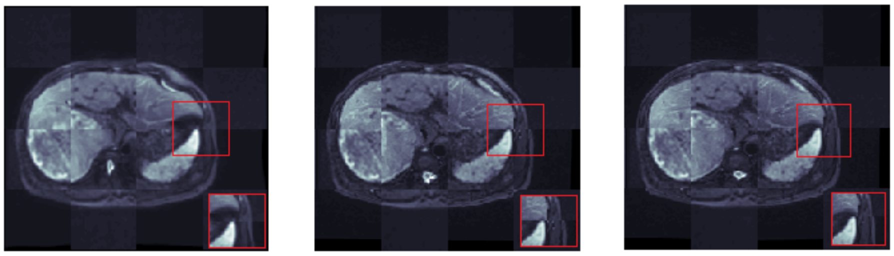
(a) The reference R (b) The template T (c) Overlay of R and T



(d) $T(\mathbf{u})$ using **New Model**, $\mathbf{GF_{er}}$=0.674, $\mathbf{MI_{er}}$=1.184, $\mathbf{NGF_{er}}$=0.8

(e) $T(\mathbf{u})$ using **NGF**, $\mathbf{GF_{er}}$=0.901, $\mathbf{MI_{er}}$=1.150, $\mathbf{NGF_{er}}$=0.856

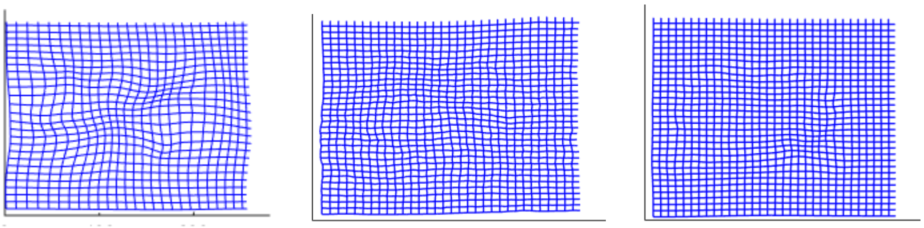
(f) $T(\mathbf{u})$ using **MI**, $\mathbf{GF_{er}}$=0.765, $\mathbf{MI_{er}}$=1.154, $\mathbf{NGF_{er}}$=0.849



(g) Overlay of R and $T(\mathbf{u})$ for new model

(h) Overlay of R and $T(\mathbf{u})$ for **NGF**

(i) Overlay of R and $T(\mathbf{u})$ for **MI**

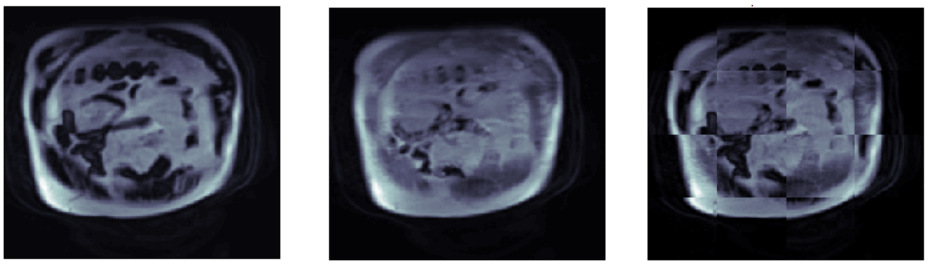


(j) $x + \mathbf{u}(x)$ using **New Model**, $\min \mathcal{C} = 0.51$

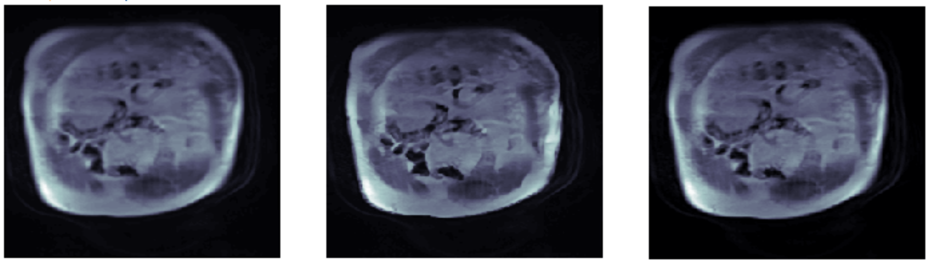
(k) $x + \mathbf{u}(x)$ using **NGF**, $\min \mathcal{C} = 0.42$

(l) $x + \mathbf{u}(x)$ using **MI**, $\min \mathcal{C} = 0.43$

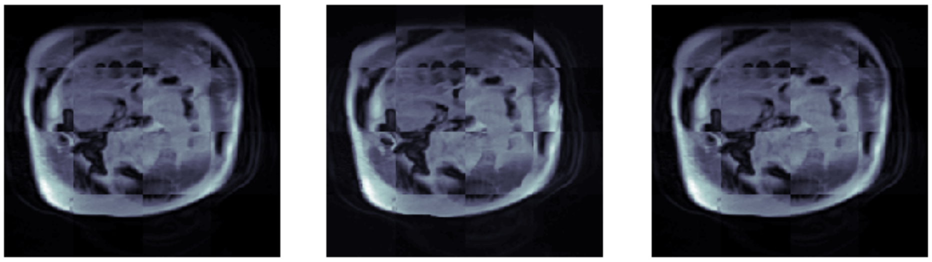
Figure 8. *Example 4: High-b- and Low-b-value Diffusion-weighted MRIs (of* 256×256*) using different models.* **New Model** *performs the best.*



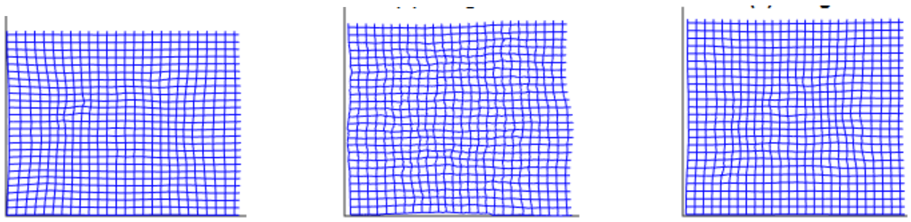
(a) The reference R (b) The template T (c) Overlay of R and T



(d) $T(\mathbf{u})$ using **New Model**, $\mathbf{GF_{er}}$=0.454, $\mathbf{MI_{er}}$=1.170, $\mathbf{NGF_{er}}$=0.623 (e) $T(\mathbf{u})$ using **NGF**, $\mathbf{GF_{er}}$=0.741, $\mathbf{MI_{er}}$=1.163, $\mathbf{NGF_{er}}$=0.656 (f) $T(\mathbf{u})$ using **MI**, $\mathbf{GF_{er}}$=0.454, $\mathbf{MI_{er}}$=1.163, $\mathbf{NGF_{er}}$=0.631



(g) Overlay of R and $T(\mathbf{u})$ for **New Model** (h) Overlay of R and $T(\mathbf{u})$ for **NGF** (i) Overlay of R and $T(\mathbf{u})$ for **MI**



(j) $x + \mathbf{u}(x)$ using **New Model**, $\min \mathcal{C} = 0.55$ (k) $x + \mathbf{u}(x)$ using **NGF**, $\min \mathcal{C} = 0.25$ (l) $x + \mathbf{u}(x)$ using **MI**, $\min \mathcal{C} = 0.95$

Figure 9. *Example 5: a pair of MRI images of higher resolution* 512×512 *by* 3 *different models.* **New Model** *and* **MI** *perform identically, both better than* **NGF**.

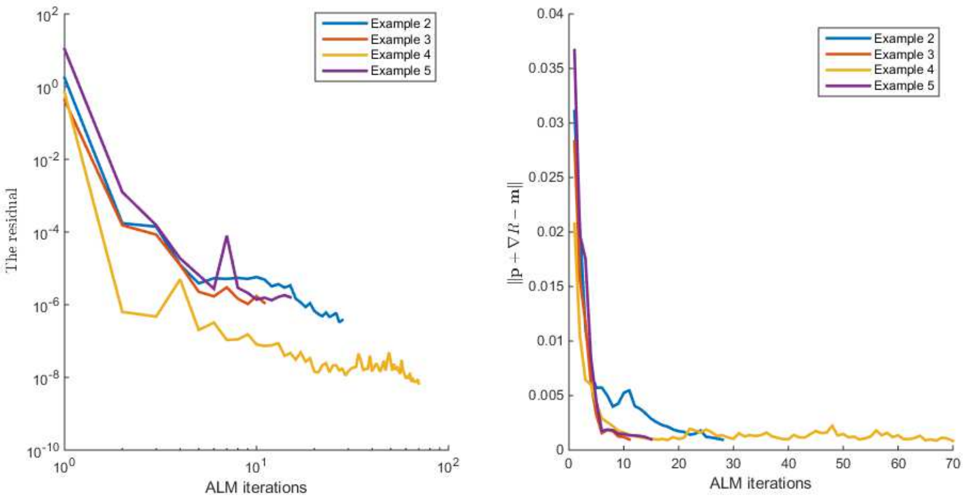


Figure 10. *Left: Log scale plot of the residual errors for* $\mathbf{u}$ *versus ALM iteration numbers for examples 2-5. Right: Plot of the error* S_{er} *values versus ALM iteration numbers for examples 2-5.*

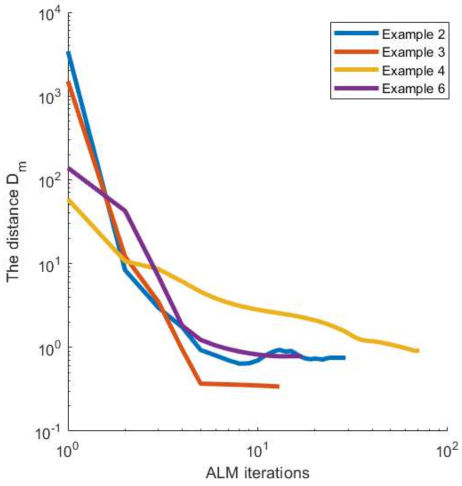
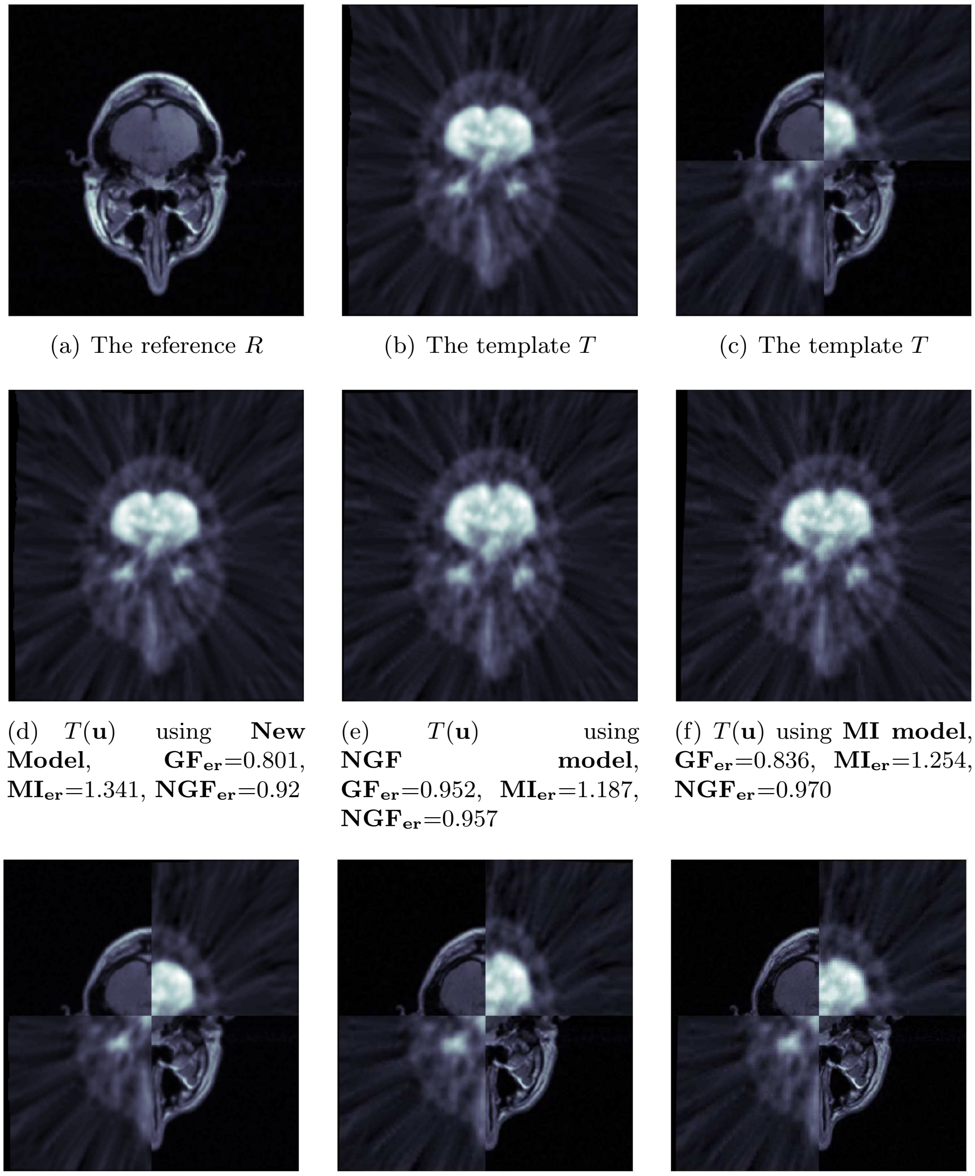


Figure 11. *Left: Log scale plot of the distance* D_m *versus ALM iteration numbers for examples 2-5.*



(a) The reference R

(b) The template T

(c) The template T

(d) $T(\mathbf{u})$ using **New Model**, $\mathbf{GF_{er}}$=0.801, $\mathbf{MI_{er}}$=1.341, $\mathbf{NGF_{er}}$=0.92

(e) $T(\mathbf{u})$ using **NGF model**, $\mathbf{GF_{er}}$=0.952, $\mathbf{MI_{er}}$=1.187, $\mathbf{NGF_{er}}$=0.957

(f) $T(\mathbf{u})$ using **MI model**, $\mathbf{GF_{er}}$=0.836, $\mathbf{MI_{er}}$=1.254, $\mathbf{NGF_{er}}$=0.970

(g) Overlay of R and $T(\mathbf{u})$ for **new model**

(h) Overlay of R and $T(\mathbf{u})$ for **NGF model**

(i) Overlay of R and $T(\mathbf{u})$ for **MI model**

Figure 12. *Example 6: Registering a PET image to an MRI vimage.* **New model** *performs better than others in this example.*