Image Registration Based on Autocorrelation of Local Structure

Zhang Li*, Dwarikanath Mahapatra, Jeroen A. W. Tielbeek, Jaap Stoker, Lucas J. van Vliet, *Member, IEEE*, and Frans M. Vos

Abstract—Registration of images in the presence of intra-image signal fluctuations is a challenging task. The definition of an appropriate *objective function* measuring the similarity between the images is crucial for accurate registration. This paper introduces an objective function that embeds local phase features derived from the monogenic signal in the modality independent neighborhood descriptor (MIND). The image similarity relies on the autocorrelation of local structure (ALOST) which has two important properties: 1) low sensitivity to space-variant intensity distortions (e.g., differences in contrast enhancement in MRI); 2) high distinctiveness for 'salient' image features such as edges. The ALOST method is quantitatively compared to the MIND approach based on three different datasets: thoracic CT images, synthetic and real abdominal MR images. The proposed method outperformed the NMI and MIND similarity measures on these three datasets. The registration of dynamic contrast enhanced and post-contrast MR images of patients with Crohn's disease led to relative contrast enhancement measures with the highest correlation ($r = 0.56$) to the Crohn's disease endoscopic index of severity.

Index Terms—Analytic signal, Hilbert transform, MIND, monogenic signal.

I. INTRODUCTION

IMAGE registration has been one of the main topics of research in the medical imaging community over the past decades. Typically, a registration method consists of three core components [1]: 1) a *deformation model* that defines how one image is transformed towards another, 2) an *objective function* that determines the similarity between two images, and 3) an

*Z. Li was with the Quantitative Imaging Group, Department of Imaging Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands. He is with the College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China (e-mail: z.li-1@tudelft.nl).

D. Mahapatra is with the Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland (e-mail: dwarikanath.mahapatra@inf.ethz.ch).

L. J. van Vliet is with the Quantitative Imaging Group, Department of Imaging Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands (e-mail: l.j.vanVliet@tudelft.NL).

J. A. W. Tielbeek and J. Stoker are with the Department of Radiology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands (e-mail: j.a.w.tielbeek @amc.uva.nl; j.stoker@amc.uva.nl).

F. M. Vos is with the Quantitative Imaging Group, Department of Imaging Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands and also with the Department of Radiology, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands (e-mail: f.m.vos@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2015.2455416

optimization strategy that searches for the values of the transformation parameters that maximize the objective function. A recent survey of the research into image registration approaches can be found in Sotiras *et al.* [1].

An appropriate objective function is pivotal to achieve accurate correspondence after registration. The choice of objective function has proven especially challenging for medical image registration. Ultimately, our focus is on registering magnetic resonance (MR) images of the gastrointestinal tract for the assessment of inflammatory bowel diseases (IBD). This registration problem is further complicated by fluctuating tissue intensity over the field of view (e.g., due to differences in contrast enhancement and the MRI bias field) as well as large deformations between corresponding structures in both images (e.g., due to respiration and peristaltic bowel movement). To be successful in registering MR images that exhibit a large intra-image signal variation, we need an objective function that is especially sensitive to the phase of the local image structures and avoid the confounding effect from the fluctuating signal intensity. We will introduce a novel criterion function based on the mean and standard deviation of the local phase derived from a multi-scale representation of its monogenic signal.

A. Related Work

Generally speaking, there are two different perspectives for looking at the objective function: *global* and *local*. From a *global* point of view, the objective function can be defined based on the statistical relationship between the image intensities. An example is Mutual Information (MI), one of the most studied registration metrics (e.g., [2], [3]). It has been widely and successfully used in both mono- and multi-modal image registration problems, e.g., [4]–[6]. However, originally being a *global* objective function, it lacks *local*, i.e., spatial, information on the local structure and cannot cope with large signal and contrast variations. Several methods were proposed addressing this problem. Initially, Pluim *et al.* [7] combined gradient information with MI in order to capture the local structure. They proposed a term, which seeks to align locations with high gradient magnitude and similar orientation. Recently, Rivaz [8] proposed the so-called SeSaMI method, which also uses gradient information. A self-similarity weighted graph-based implementation is introduced using α-Mutual Information (α-MI) for nonrigid image registration. The method employs a self-similarity measure based on local structural information, which is invariant to rotation and to local affine intensity distortions. However, the method comes with a rather high computation cost because multi-dimensional features

are involved to build SeSaMI. Alternatively, Studholme *et al.* [9] introduced a regional label as an extra "channel" to the intensity joint histogram to compute the so-called Regional Mutual Information (RMI). The RMI was shown robust to local contrast variations over diseased brain tissue. Recently, Loeckx *et al.* [10] proposed a measure called Conditional Mutual Information (CMI) which extends the intensity joint histogram with a third channel containing spatial information. Although these methods introduce local information into a *global* objective function, it has been noticed [11] that finding an accurate correspondence remains difficult, especially due to the many local optima that generally accompany most non-rigid deformation models.

From a *local* point of view, the objective function is usually defined based on a sum of point-to-point or (small) region-to-region correspondences between images. Sum of squared differences (SSD) and cross correlation (CC) are exemplary methods to do so. However, they are only reliable under the assumption that the intensity relationship between the images is linear. For many registration problems, especially concerning mono-modal images with space-variant intensity distortions and multi-modal images, this assumption is not correct.

Several methods [12]–[14] were proposed for tackling this non-linear intensity relationship by relying on features based on local orientation rather than intensities. The first two methods [12], [13] use the gradient orientation in a multi-scale fashion, but restricted to voxels with a high gradient magnitude. The third method [14] employs the local orientation as obtained by the local structure tensor (LST) [15], [16] . Here, the eigenvector corresponding to the largest eigenvalue of the structure tensor points in the dominant orientation of the local signal. This feature was shown to yield a more consistent registration for small image features (e.g., thin lines). A limitation of these methods is that the local orientation is ill-defined in areas where two or more unimodal oriented patterns come together (e.g., bifurcations of vessels).

One might also consider the local entropy as measures of the local image structure. Following up on that idea, local entropy was introduced [17]. The intensity histogram was first calculated over a small patch centered at each voxel in the image. Then, the Shannon entropy for each voxel was calculated over the smoothed histograms in which discretization errors were suppressed by use of a Parzen kernel. The assumption behind using the entropy image was that intensity variations would occur at corresponding locations in the different images. Accordingly, the SSD of entropy images was used as the objective function. However, this method is sensitive to noise when calculating the histogram over small patches.

Alternatively, local phase and local amplitude from the monogenic signal were proposed to represent local image structure [18], [19] . The local amplitude of the monogenic signal represents the local (square-root of the) energy of the signal and the local phase is a signature determining the balance between the even and the odd components of the local signal. Image registration steered by similarity in local phase and local amplitude was introduced [20]–[22]. However, these registration methods used a voxel-wise representation of structure, without particularly emphasizing the structural resemblance in the surrounding

neighborhood. Another drawback of merely using a voxel-wise phase representation is the sensitivity to noise. It was already mentioned in [20] that noise can only be dealt with by carefully choosing certain parameters.

Recently, a local image descriptor based on neighborhood information called MIND (modality independent neighborhood descriptor) was proposed [11]. The method was derived from the concept of self-similarity, which was first proposed by Buades *et al.* [23]. MIND embeds the local structure in a vector representation for each voxel. This descriptor essentially comprises the local autocorrelation function (see also Section II-C). It has resulted in better registration results compared to other, state-of-the-art methods. MIND was reported to be a "distinctive" descriptor, which might be important for registering images with many degrees of freedom [11]. One may notice, however, that small image patches by themselves are not a good representation of the local structure. Such patches capture not only true structure information originating from signal differences across organ boundaries, but also undesired information. For instance, differences in contrast enhancement are 'uncharacteristic'. MIND does not fully suppress such influences. Fortunately, one can use a band-pass filter to calculate the local phase for image patches. The phase information primarily describes the desired structure that is related to sharp edges. Complimentary, the contrast difference is contained in the local amplitude. By neglecting the local amplitude information, the information left (local phase) is insensitive to differences in contrast enhancement. Furthermore, the recently described self-similarity context (SCC) descriptor [24] might improve over the MIND approach for specific applications, e.g., registering ultrasound (US) to MRI. The SSC method enhances MIND in terms of robustness to noise. This is because SSC calculates all pairwise self-similarities of patches within a six-neighborhood rather than only the self-similarities to the center patch as MIND does. Observe that MIND and SSC are very similar and rely on the same self-similarity measure. As such SSC is also not robust to contrast variations.

B. Objective

In this paper we present a novel measure of the local image structure based on the monogenic signal, which is insensitive to "uncharacteristic" information such as differences in contrast enhancement and the MRI bias field. Subsequently, we integrate this measure in the aforementioned MIND principle. As such, we exploit the autocorrelation of local structural information (ALOST) rather than the autocorrelation of intensity (MIND). We demonstrate the benefit of image registration using ALOST by applying it to images that have space-variant intensity distortions.

II. METHOD

The ALOST method is designed to register images distorted by contrast variations, e.g., by the MRI bias field or fluctuating gadolinium uptake. The method involves several image processing techniques. The monogenic signal [18] is introduced to measure the local image structure of 3D images. The monogenic signal is an extension of the analytic signal to higher dimensions (Section II-A). It is calculated using specific band-pass filters.

Two complementary features are derived from the monogenic signal: local phase and phase congruency (Section II-B). We integrate these features in a patch-based representation called ALOST (Section II-C). Finally, the image registration pipeline is described (Section II.D).

A. The Analytic and Monogenic Signal

The *analytic representation* of a signal is based on the concept that negative frequency components of a real-valued signal are essentially superfluous due to the Hermitian symmetry of its Fourier spectrum. Specifically, if $f(x)$ is a real-valued signal and $\mathcal{F}(\omega)$ its Fourier transform, then the Fourier spectrum

$$\mathcal{F}_A(\omega) = \mathcal{F}(\omega)(1 + \text{sgn}(\omega)) \tag{1}$$

contains only positive frequency components and is reversible into $\mathcal{F}_{(\omega)}$. The inverse Fourier transform of $\mathcal{F}_A(\omega)$ is called the analytic signal [15]

$$f_A(x) = f(x) - i f_H(x), \tag{2}$$

where $f_H(x)$ is Hilbert transform of $f(x)$:

$$f_H(x) = H(f(x)) = \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} \frac{f(\tau)}{\tau - x} d\tau. \tag{3}$$

Without loss of information the negative frequency components of a real-valued signal are discarded, but instead one has to deal with a complex-valued signal. Incidentally, notice that the conversion from the complex-valued analytic signal back to the underlying real-valued signal is just a matter of discarding the imaginary part from (2). The analytic signal representation makes certain attributes of the signal more accessible. Typically, the analytic signal is described by the local amplitude $|f_A(x)|$ and local phase $\varphi(x)$ defined as:

$$|f_A(x)| = \sqrt{f_A(x) f_A^*(x)} = \sqrt{f^2(x) + f_H^2(x)}, \tag{4}$$

$$\varphi(x) \equiv \arctan\left(\frac{f_H(x)}{f(x)}\right). \tag{5}$$

Generally, direct calculation of these quantities cannot be done, because the Hilbert transform requires an infinite support filter, see (3). Therefore, the local amplitude and phase are conventionally estimated based on a band-pass filtered version of the signal. If $q_e(x)$ represents the band-pass filter, we have that the analytic signal is approximated by:

$$\begin{aligned} \hat{f}_A(x) &= f_A(x)^* q_e(x) = [f(x) - i f_H(x)]^* q_e(x) \\ &= f(x)^* q_e(x) - i H(f(x)^* q_e(x)) = f_e(x) - i f_o(x), \\ &= [q_e(x) - i q_0(x)]^* f(x) \end{aligned} \tag{6}$$

where $q_o(x)$ is the Hilbert transform of $q_e(x)$, so that these filters form a quadrature-pair. In particular, we choose $q_e(x)$ to be an even, real-valued, band-pass filter (hence the subscript e) with zero DC-value to be invariant to an arbitrary chosen amplitude offset. Furthermore, its real-valuedness and even symmetry implies that there is no change of phase information if we choose $q_e(x)$ such that its Fourier transform satisfies $Q_e(x) \geq 0$. In

turn, $q_0(x)$ is odd. Observe that the response of the signal to $q_e(x)$ gives the local real-and-even component of the approximated analytic signal and that the response to $-q_0(x)$ yields the local imaginary-and-odd component of it.

The monogenic signal [18] extends the analytic signal for one-dimensional signals to arbitrary multi-dimensional signals by means of the Riesz-transform. We restrict ourselves to 3D images for illustration purposes. The monogenic signal relies on a rotationally symmetric, real-valued, zero-mean band-pass filter whose Fourier transform is non-negative to avoid phase flipping. This filter is used to extract the even component of the signal:

$$f_e(\mathbf{x}) = f(\mathbf{x})^* q_e(\mathbf{x}), \tag{7}$$

where $f(\mathbf{x})$ is a 3-D image and $q_e(\mathbf{x})$ is an isotropic band-pass filter [18]. The choice of the band-pass filter is not a trivial task. We will discuss it in the next section.

Three filters are constructed to extract the three odd components (i.e., one per image dimension) from the signal by applying the 3-D Riesz transform to the isotropic band-pass filter. The frequency responses of the filters are given in the 3D Fourier domain by:

$$\mathbf{Q}_R(\mathbf{u}) = \mathbf{Q}_o(\mathbf{u}) \frac{i\mathbf{u}}{|\mathbf{u}|} Q_e(\mathbf{u}), \tag{8}$$

in which $Q_e(\mathbf{u})$ represents the Fourier transform of $q_e(\mathbf{x})$. Notice that the term $(\mathbf{u}/|\mathbf{u}|)$ yields the same frequency response along its primary axes as the 1-D Hilbert transform. Also, while $\mathbf{u}$ represents a basis vector of the Fourier space, there are three odd components for a 3D image. The response to (8) yields the Riesz transform of $f(\mathbf{x})$ defined as:

$$\mathbf{f}_R(\mathbf{x}) = \mathcal{F}^{-1}\{\mathcal{F}\{f_e(\mathbf{x})\} \cdot \mathbf{Q}_R\}. \tag{9}$$

Finally, the monogenic signal (expressed in quaternions) is:

$$f_M(\mathbf{x}) = f_e(\mathbf{x}) - (i, j, k)\mathbf{f}_R(\mathbf{x}). \tag{10}$$

The local amplitude and phase of the monogenic signal are defined as:

$$A(\boldsymbol{x}) = |f_M(\mathbf{x})| = \sqrt{f_e(\mathbf{x})^2 + |\mathbf{f}_R(\mathbf{x})|^2}, \tag{11}$$

and

$$\varphi(\mathbf{x}) = \text{atan}\left(\frac{|\mathbf{f}_R(\mathbf{x})|}{f_e(\mathbf{x})}\right). \tag{12}$$

Notice that with this definition the phase $\varphi(\mathbf{x})$ ranges from $-\pi/2$ to $\pi/2$. Although the orientation of the phase can be expressed in two angles, these are not used in the proposed registration method.

B. The Band-Pass Filter and Representation of Local Structure

The choice of band-pass filter is important in obtaining some key properties [25]. We choose a log-Gabor filter which is essentially a shifted Gaussian function on a logarithmic scale in the frequency domain. We opt to do so because it is a zero DC filter with a tunable bandwidth. Furthermore, log-Gabor filters are invariant to an additive polynomial function of order $n < \infty$.

The filter $q_e(\mathbf{x})$ is defined in the frequency domain as:

$$Q_e(\boldsymbol{\omega}) = n_c \exp\left(\frac{-\ln^2\left(\frac{\omega}{\omega_0}\right)}{2\ln^2(k_\beta)}\right), \qquad (13)$$

where n_c is a normalization constant, $\boldsymbol{\omega_0}$ the central (tuning) frequency and $k_\beta (0 < k_\beta < 1)$ a parameter related to the bandwidth β of the filter:

$$\beta = -\frac{2\sqrt{2}}{\sqrt{\ln 2}}\ln(k_\beta). \qquad (14)$$

In our experiments the log-Gabor filter was applied to four scales ($\boldsymbol{\omega_0} = 3, 9, 27$ and 81) [22]. Using $k_\beta = 0.47$ yields a log-Gabor filter with 2.6 octaves in bandwidth which is the smallest bandwidth of edge operators without significant aliasing [25].

At each scale, we compute the local amplitude $A_i(\mathbf{x})$ and local phase $\varphi_i(\mathbf{x})$ using (11) and (12). The local energy model postulates that features are perceived at points in an image where the Fourier components are in phase [26]. As such, congruency of phase at any angle produces a clearly perceived feature [27]. The angle at which the congruency occurs dictates the feature type, for example, a step corresponds to $\varphi = 0$ and a peak to $\varphi = 1/2\pi$.

Initially, a measure for phase congruency (PC) was developed by Morone and Owens [26]:

$$PC_1(\mathbf{x}) = \frac{|E(\mathbf{x})|}{\sum_{i=1}^n A_i(\mathbf{x})}, \qquad (15)$$

where $E(\mathbf{x})$ is defined as the scale accumulated monogenic signal:

$$E(\mathbf{x}) = \sum_{i=1}^n f_{M,i}(\mathbf{x}), \qquad (16)$$

and (for our 3D image space), $|E(\mathbf{x})|$ is given by:

$$|E(\mathbf{x})| = \left[\left(\sum_{i=1}^n f_{e,i}(\mathbf{x})\right)^2 + \left(\sum_{i=1}^n f_{R1,i}(\mathbf{x})\right)^2 \right.$$
$$\left. + \left(\sum_{i=1}^n f_{R2,i}(\mathbf{x})\right)^2 + \left(\sum_{i=1}^n f_{R3,i}(\mathbf{x})\right)^2\right]^{1/2}, \quad (17)$$

with A_i the local amplitude (11) and $f_{R1,i}$, $f_{R2,i}$ and $f_{R3,i}$ the three components originating from Riesz transform $\mathbf{f_{R,i}}(\mathbf{x})$ (9) at scale i. $PC_1(\mathbf{x}) = \sum_{i=1}^n A_i(\mathbf{x})\cos(\varphi_i(\mathbf{x}) - \overline{\varphi}(\mathbf{x}))/\sum_{i=1}^n A_i(\mathbf{x})$ can be written as a weighted sum of cosines, where the $\overline{\varphi}(\mathbf{x})$ is the phase of $E(\mathbf{x})$ henceforth referred to as the mean phase:

$$\overline{\varphi}(\mathbf{x}) = \arctan\left(\frac{\sum_{i=1}^n |\mathbf{f}_{R,i}(\mathbf{x})|}{\sum_{i=1}^n f_{e,i}(\mathbf{x})}\right). \qquad (18)$$

If all scale components are in phase, the complex vectors would be aligned and the phase congruency would be 1. Alternatively, if there is no coherence of phase, the phase congruency becomes zero. Phase congruency provides a measure that
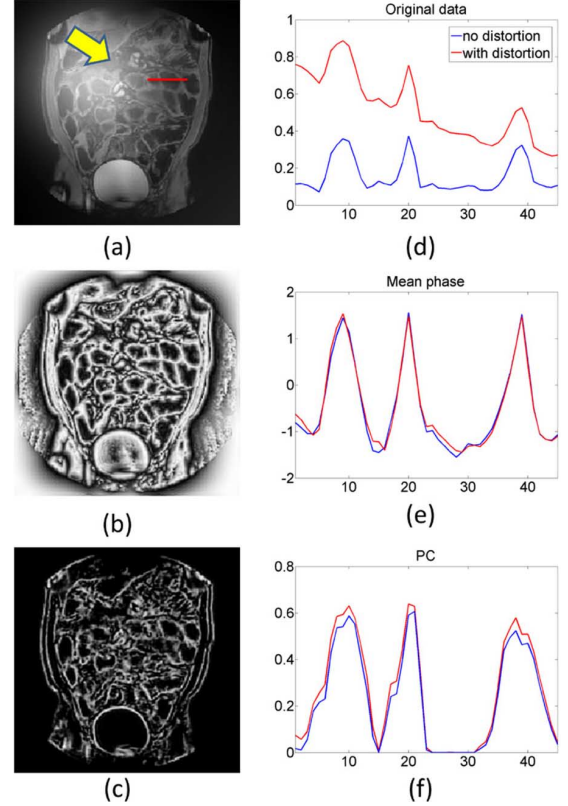

Fig. 1. (a) An abdominal MR image distorted by a simulated space-variant intensity distortion; (b,c) mean phase image and phase congruency images extracted from (a); (d)–(f): intensity profiles along the red lines in (a), (b) and (c) from the distorted image and undistorted image respectively.

is independent of the magnitude, making it invariant to global trends in the signal and variations in contrast. Moreover, averaging over different scales decreases the influence of noise on the estimated phase.

Later, Kovesi [28] developed a modified phase congruency measure, which subtracted the sine of the phase deviation to produce a more localized response:

$$PC_2(\mathbf{x}) =$$
$$\frac{\sum_{i=1}^n W(\mathbf{x})\lfloor A_i(\mathbf{x})[\cos(\varphi_i(\mathbf{x})-\overline{\varphi}(\mathbf{x}))-|\sin(\varphi_i(\mathbf{x})-\overline{\varphi}(\mathbf{x}))|] - T\rfloor}{\sum_{i=1}^n A_i(\mathbf{x}) + \varepsilon},$$
$$(19)$$

where the factor $W(\mathbf{x})$ reflects the spread in frequency. T is a threshold ensuring that only values exceeding the noise floor are taken into account. Therefore, the symbols $\lfloor \ \rfloor$ denote that the quantity enclosed is equal to itself when larger than T and zero otherwise. Finally, ε is a small positive constant to avoid division by zero. More details can be found in the work by Kovesi [28].

Fig. 1 shows the mean phase and phase congruency (PC_2) applied to an image with a simulated, spatially-varying intensity distortion [29]. The red profiles show that the mean phase is not sensitive to the applied signal distortion and that the phase congruency is high for both edges and peaks. The blue profiles show the results from the undistorted images for comparison. It can be seen that the profiles from the distorted image and undistorted image show a strong resemblance, both for the mean phase and phase congruency. Notice also that particularly

larger features exhibiting a sharp intensity transition such as the bladder boundary yield a high phase congruency, whereas smaller features e.g., thin bowel wall have a low phase congruency.

C. Autocorrelation of Local Structure

The mean phase and phase congruency are voxel-wise descriptors of the local structure that are not sensitive to global trends. They have also limited sensitivity to noise if one chooses the band-pass filters carefully (i.e., by applying appropriate cut-off frequencies). Indeed, a mere phase based method can still suffer from noise. Neither mean phase nor phase congruency by itself captures the coherence of structural information in a neighborhood. A patch-based method [11], [24] solves this issue and simultaneously suppresses the sensitivity to noise even further. One way to do so is using a local autocorrelation (AC) function which is embedded in the local sum of squared differences (SSD) between two patches of size P in image I:

$$D_P(I, \mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{p} \in P} \left(I(\mathbf{x}_1 + \mathbf{p}) - I(\mathbf{x}_2 + \mathbf{p}) \right)^2$$

$$= \sum_{\mathbf{p} \in P} I^2(\mathbf{x}_1 + \mathbf{p}) - 2 \underbrace{\sum_{\mathbf{p} \in P} I(\mathbf{x}_1 + \mathbf{p}) I(\mathbf{x}_2 + \mathbf{p})}_{AC_P(I, \mathbf{x}_1, \mathbf{x}_2)}$$

$$+ \sum_{\mathbf{p} \in P} I^2(\mathbf{x}_2 + \mathbf{p}), \quad (20)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the centers of the two image patches P of size $D \times D \times D$ (in 3-D case). The term D_p is also known as the local self-similarity. D_p gives a small value when the local autocorrelation function AC_p yields a high value. As such the modality independent neighborhood descriptor (MIND) [11] incorporates the AC_p as follows:

$$\text{MIND}(I, \mathbf{x}, \mathbf{r}) = \frac{1}{n} \exp\left(\frac{-D_p(I, \mathbf{x}, \mathbf{x} + \mathbf{r})}{V(I, \mathbf{x})} \right), \quad (21)$$

with $\mathbf{r}$ an offset in a predefined neighborhood $\mathbf{R}$ of size $R \times R \times R$ around position $\mathbf{x}$, n a normalization constant (chosen such that the maximum value of MIND at position $\mathbf{x}$ is 1) and $V(1, \mathbf{x})$ represents the average SSD in a small neighborhood N around $\mathbf{x}$:

$$V(I, \mathbf{x}) = \frac{1}{\text{size}(N)} \sum_{\mathbf{n} \in N} D_P(I, \mathbf{x}, \mathbf{x} + \mathbf{n}). \quad (22)$$

In other words, MIND returns an 1-D vector of length $R^3 - 1$ which represents all correlations between a neighborhood P centered at voxel $(\mathbf{x})$ and a neighborhood P centered at a voxel around it $(\mathbf{x} + \mathbf{r})$. Three parameters have to be defined in MIND: the patch size D, the search neighborhood size R, and the size of the variance calculation neighborhood N. We followed the same setup as [11]: $N = N_6$ which means a six-connected neighborhood and patch size $D = 3$. Only the size of the search region R is a free parameter and was trained for different datasets. More details for choosing R can be found in Section III.

The original MIND descriptor is applied to the image intensity itself, which makes it sensitive to global trends (e.g., intensity distortions, see Fig. 1(a)). We propose to combine the best of two worlds: 1) the structural measures of mean phase
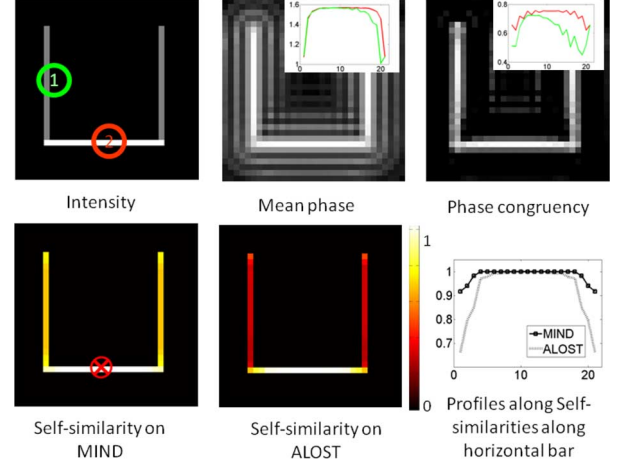

Fig. 2. A "toy" example demonstrating the high distinctiveness of the ALOST descriptor. The first row shows the intensity and corresponding mean phase and phase congruency images. The insets in the mean phase and phase congruency images represent the signal profile along the vertical bar (top-to-bottom in green) and the horizontal bar line (red). The second row shows the self-similarity of the MIND and ALOST descriptors with respect to the red cross and signal profiles from both self-similarity images.

and phase congruency and 2) the neighborhood description of MIND. The actual computation of MIND (as well as the new concept) is performed through the sum of square differences. The above derivation explains, however, that the autocorrelation of intensity is underlying MIND. Accordingly, we exploit the autocorrelation of local structure (ALOST). ALOST describes the local structure for each image voxel as a $2(R^3 - 1)$-dimensional vector:

$$ALOST(I, \mathbf{x}, \mathbf{r}) =$$
$$[\text{MIND}(\overline{\varphi}(I), \mathbf{x}, \mathbf{r}), \text{MIND}(\text{PC}(I), \mathbf{x}, \mathbf{r})], \quad (23)$$

where $\overline{\varphi}(\mathbf{x})$ and $\text{PC}(\mathbf{x})$ are the mean phase and phase congruency from, respectively, (18) and (19). Thus, the phase information over all scales is represented by a mean value, encoding for the type of structure (e.g., ridge, slope, etc.), and a value reflecting the spread of the scale-dependent phases around the mean phase, signifying the reliability of such a feature. It will be shown that the proposed ALOST descriptor enhances the distinctiveness of image features over separate representations based on either mean phase or phase congruency. In order to visualize the distinctiveness of the descriptors we measure the self-similarity of the local structure descriptors as:

$$\text{SELFSIM}(DESCRIPTOR, I, \mathbf{x}, \mathbf{r}, \mathbf{m})$$
$$= 1 - \frac{1}{|M|} |DESCRIPTOR(I, \mathbf{x}, \mathbf{r})$$
$$- DESCRIPTOR(I, \mathbf{x} + \mathbf{m}, \mathbf{r})|_1, \quad (24)$$

where *DESCRIPTOR* is either MIND or ALOST, input image I either the intensity, the mean phase or the phase congruency and M the region size over which the self-similarity is measured. This gives rise to four descriptors denoted as ALOST, MIND_I, MIND_MP, MIND_PC.

Fig. 2 shows a toy example in which the self-similarities of MIND and ALOST are compared. The top left image contains an artificial input image consisting of three bars: the intensity of the horizontal bar is twice the intensity of the vertical bars.
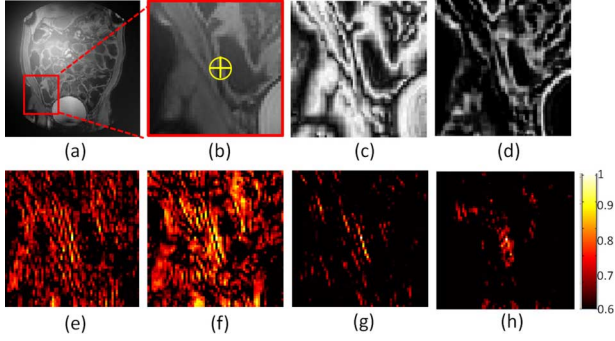
Fig. 3. (a) An abdominal MR image distorted by a simulated space-variant intensity distortion (a) and a region-of-interest (b); (c, d) mean phase and phase congruency of (b); (e)-(h), the self-similarity map with respect to the yellow cross based on respectively MIND_I , MIND_MP, MIND_PC and ALOST applied to the region-of-interest.

The top middle and right images display mean phase and phase congruency of the input image. The insets plot the signal along the horizontal bar (red, from left to right) and the vertical bar (green, from top to bottom). The bottom-left and middle images show the self-similarity of MIND on intensity and ALOST around the indicated point. Finally, the bottom-right plot shows the self-similarity profile along the horizontal bars for the MIND and ALOST descriptors.

The inset in the top-middle image shows that the mean phase remains largely constant, but dips slightly faster when approaching the corner compared to the line end (see the green line). The inset in the top-right image conveys that the phase congruency is more sensitive to a different structure in the surroundings than the mean phase. Furthermore, it can be observed that in both images the signal starts to decrease further away from the corner along the grey bar than along the white bar, see the right part of the insets. This is due to the difference in intensity: mixing a larger intensity into the filter kernels affects the output more than mixing a lower intensity.

It can be observed from the bottom-right image that the self-similarity of ALOST decreases before MIND does. This confirms that ALOST comes with increased distinctiveness. The example shows that this is especially due to the sensitivity of the phase congruency.

Fig. 3 shows the self-similarity of the various descriptors in a region-of-interest from an abdominal image corrupted by a simulated intensity distortion. Clearly, Fig. 3(h) shows the "distinctiveness" of ALOST around the voxel that we picked. Note that MIND applied to the mean phase shows the least distinctiveness among all descriptors.

D. Image Registration Pipeline

We implemented the registration pipeline as used in MIND [11]. The objective function which is going to be minimized consists of a data term and a regularization term [30]:

$$\mathbf{u}_m^* = \arg\min_{\mathbf{u}_m} \left(\sum_x \left(\frac{1}{|R|} \sum_{\mathbf{r} \in R} |\text{ALOST}(\boldsymbol{I_f}, \mathbf{x}, \mathbf{r}) \right. \right.$$

$$\left. \left. -\text{ALOST}(\boldsymbol{I_m}, \mathbf{x}+\mathbf{u}_m, \mathbf{r})| \right)^2 + \alpha \|\nabla \mathbf{u}_m(\mathbf{x})\|^2 \right), \quad (25)$$

where $\mathbf{u}_{\mathrm{m}} = (u, v, w)$ is the deformation field for each voxel in the moving image and $\mathbf{u}_m^*$ the optimal transformation. The data term of the objective function is defined as the sum of absolute descriptor differences. The second term performs local diffusion regularization. The objective function is then optimized by means of the Gaussian-Newton method. While doing so a multi-resolution procedure was followed involving three levels corresponding to down-sampling by a factor of 4, 2 and 1 respectively (1 refer to original resolution). The weighting parameter α was set to 0.1 for all the experiments (We varied the alpha from 0.01 to 0.1, which did not give any significant difference). More implementation details can be found in [11].

III. RESULT AND DISCUSSION

The four descriptors (MIND_I, MIND_MP, MIND_PC, and ALOST) were evaluated on several datasets: abdominal MRI data with synthetic spatially-variant intensity distortions and synthetic spatial deformations (Section III-A), thoracic CT images (Section III-B), and abdominal MR images (Section III-C). The computation time for each descriptor depends linearly on the number of voxels. The average registration time for the largest dataset (thoracic images of size $256 \times 256 \times 94$) was less than 10 minutes on a personal computer equipped with an Intel® Core™2 Quad Processor Q8400 clocked at 2.66 GHz and 4GB RAM memory.

A. Synthetic Data

We simulated a corrupted image by applying an intensity distortion to abdominal MR image data (400×400 pixels with a resolution of 1×1 mm^2) as described by Myronenko et al. [29]. Firstly, a spatially-varying intensity distortion was applied to an abdominal MR image to create the fixed image (see Fig. 4(a)). Second, another intensity distortion was applied to the same (initial) abdominal image in a comparable manner. Thereafter, a geometric distortion with a thin-plate spline (TPS) model was applied to this image to get a second image, used as the moving image (see Fig. 4(a)). The grid size of the thin-plate spline model was $11 * 11$ mm^2 and to the node displacements a Gaussian distributed random perturbation was added (zero mean and 5 mm stdev). The maximal displacement was approximately 10 mm, which we consider a large local deformation. The intensity distorted regions are contained in the dashed red box of the moving image. Subsequently, we used the Normalized Mutual Information (NMI) [5] as a baseline metric as well as the four approaches described in Section II-C to register these data: MIND_I, MIND_MP, MIND_PC and ALOST (see the first row of Figs. 4(b)–4(f). Enlarged versions of the regions corresponding to the aforementioned dashed red boxes after registration are shown in respectively the second row of Figs. 4(b)–4(f). The fixed image is displayed in green and the registered moving image in magenta. A perfect registration yields a grey-scale image. Any color denotes misregistration. For display purposes, the color-coding was applied to the images without intensity distortion.

It can be seen that MIND_MP, MIND_PC and ALOST gave better registration results in the region with intensity distortions than NMI and MIND_I. Particularly, one can see from Figs. 4(b) and 4(c) that the registration is off in the regions
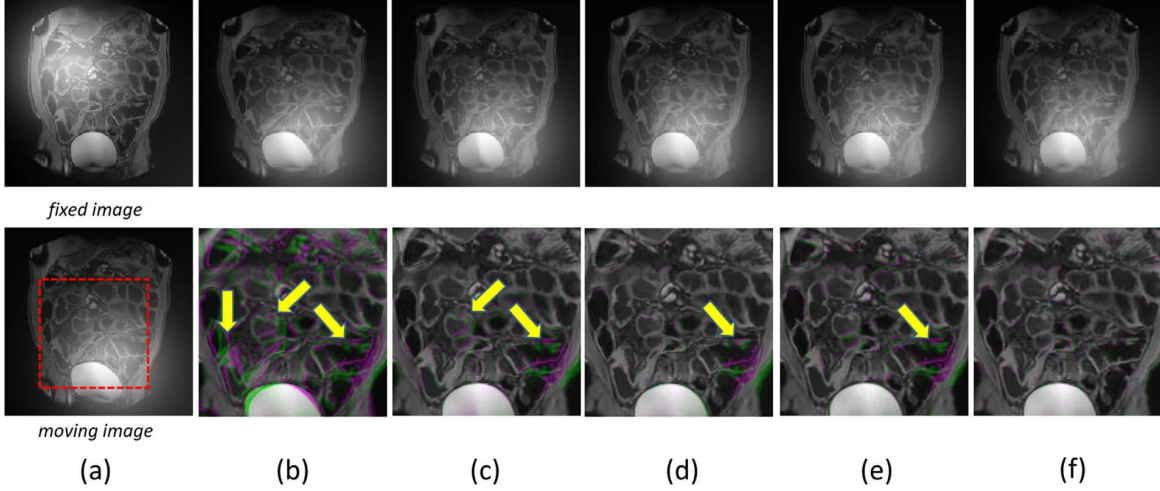
Fig. 4. (a) An abdominal MR image with a simulated space-variant intensity distortion (fixed image) and the same abdominal MR image affected by a different simulated intensity distortion and a spatial deformation (moving image); (b)-(f) the top images show the moving image after registration based on respectively NMI, MIND_I, MIND_MP, MIND_PC and ALOST; the bottom image contain the fixed image in the green channel and the registered, moving image in the red and blue channels: proper registration gives grey tone and misregistration results in green or magenta tones; the yellow arrows indicate misregistration in the regions with large contrast variation and spatial deformation in the moving image.
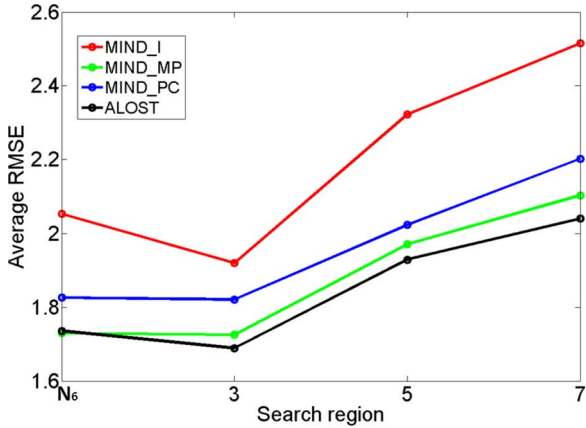

Fig. 5. The average root mean square error (RMSE) (mm) of the estimated deformation fields as a function of the search region size (R) on synthetic abdominal image training data (20 samples). R = N4 corresponds to a four-neighborhood region, R = 3 to a 3×3 region, R = 5 to a 5×5 region and R = 7 to a 7×7 region.

pointed out by the yellow arrows. This illustrates that NMI and MIND_I are sensitive to intensity fluctuations. Furthermore, ALOST yielded better results than registration only based on MIND_MP or MIND_PC in the areas again denoted by the yellow arrows. Registration based on the MIND_MP descriptor particularly failed in a region that was subject to a large deformation (Fig. 4(d)) as MIND_MP is not sufficiently distinctive (see Fig. 3(c)) to register this part. Registration based on the MIND_PC descriptor also caused misregistration in the region indicated in Fig. 4(e). This happened because MIND_PC merely emphasizes 'larger' features with a strong intensity transition (see Fig. 1(c)) and does not perform well in complex regions in which small details are relevant. Such regions only give a reliable phase response on the smallest scale. Essentially, ALOST combines the merits of MIND_MP and MIND_PC. Therefore, it facilitates enhanced distinction and is insensitive to intensity distortions. This experiment shows that ALOST has the potential to register images suffering from large spatial deformations and strong intensity distortions.

We generated 100 synthetic abdominal MR images with a random spatial deformation and a random intensity distortion (as above) to quantitatively compare our method to the other techniques. The NMI method was taken from the *elastix* toobox [31] using a B-spline transform with a histogram bins setting of 64 (We tested three settings for the number of histogram bins: 64, 128 and 256; we did not observe any significant differences). This setting was used in all the following experiments. The other four methods were tested using the MIND_I registration framework, which needs careful tuning of the search region $\mathbf{R}$. Therefore, we first used 20 pairs of synthetic images to identify the best $\mathbf{R}$ for each method. The average root mean square error (RMSE) of the estimated deformation field with each method for varying $\mathbf{R}$ is shown in Fig. 5. Apparently, all four methods gave the lowest average RMSE at $\mathbf{R} = 3$. Notice also that the ALOST method gave the lowest average RMSE for all settings. We subsequently used $\mathbf{R} = 3$ to test the methods on the remaining 80 datasets. The results are contained in Table I. One can see that the phase-based methods MIND_MP, MIND_PC and ALOST gave better results than NMI and MIND. This is because these three methods are insensitive to the intensity distortion. One can also see that the ALOST method gave the lowest average RMSE among the phase-based methods because of the better distinctiveness. A two-sided signed rank test demonstrated that ALOST had a significantly better outcome than all other methods ($p < 0.05$).

B. Registration of Thoracic CT Images

The MIND_I method was evaluated on publicly available 4D thoracic CT scans of ten patients [32]. Each scan has 300 anatomical landmarks annotated by thoracic imaging experts in both the inhale and exhale scans. The inter-observer error is less than 1 mm. The registration of inhale and exhale scans can be seen as a mono-modal registration problem. However, due to gas density differences between inhale and exhale scans, an objective function based on minimizing the sum of squared differences failed in several cases [11].

TABLE I
AVERAGE ROOT MEAN SQUARED ERROR (RMSE) OF THE ESTIMATED DEFORMATION FIELD PRIOR TO (INITIALLY) AND AFTER REGISTRATION USING NMI, MIND_I, MIND_MP, MIND_PC AND ALOST ON SYNTHETIC ABDOMINAL MR IMAGES. THE NUMBER PRINTED IN BOLD FACE IS THE MINIMUM (BEST RESULT). THE STANDARD DERIVATION OF THE RMSE IS GIVEN BETWEEN BRACKETS

	Initially	NMI	MIND_I	MIND_MP	MIND_PC	ALOST
mRMSE	3.43 (0.29)	4.61 (1.02)	1.97 (0.27)	1.83 (0.30)	1.85 (0.24)	**1.73 (0.24)**

TABLE II
MEAN TARGET REGISTRATION ERROR (MTRE) OF 300 LANDMARKS PRIOR TO REGISTRATION (I.E., INITIALLY), AND AFTER REGISTRATION USING NMI, MIND_I, MIND_MP, MIND_PC AND ALOST ON 10 DATASETS MADE AVAILABLE BY DIR-LAB. THE NUMBER PRINTED IN BOLD FACE IS THE MINIMUM (BEST RESULT)

	Initially	NMI	MIND_I	MIND_MP	MIND_PC	ALOST
MTRE	8.46 (3.33)	2.90 (1.55)	2.00 (0.96)	2.26 (1.10)	2.46 (1.51)	**1.89 (0.89)**

We again evaluated five methods (NMI, MIND_I, MIND_MP, MIND_PC, ALOST) on these datasets by calculating the mean target registration error (MTRE) of the landmarks after registering the exhale to the inhale scans (see Table II). While doing so, we used $\mathbf{R} = N_6$ (i.e., a six-neighborhood setting) in all four methods. Note that this setting was reported in [11] to yield the best registration result for MIND_I. What is more, we empirically found that MIND_MP, MIND_PC and ALOST also gave the best registration result with $\mathbf{R} = N_6$. One should notice that our registration method is not particularly designed for registration of the complete thorax. For instance, we do not consider the sliding motion to the side of the lungs as [33] does. However, the annotations are more in the central part of the lungs, which our method should be able to register accurately.

Notice that NMI gave the worst registration results compared to the other methods. In this dataset, the gas density changes due to compression introduce contrast changes between tissue and air [32]. NMI is not suited to deal with this problem as was previously noticed in [11]. Furthermore, we found that MIND_PC gave the second worst result among the five methods. Again this is because important features such as the tubes of the airways are only a few pixels wide. Consequently, the phase congruency was low on these details. Since the aforementioned tubes are invisible on the larger scales, the phase is uncorrelated across scales. Consequently, the registration in MIND_PC will be dominated by the larger structures and yield an ambiguous outcome on details. Finally, a two-sided signed rank test was applied to the MTRE showing that ALOST had a significantly better outcome for all but cases 1, 7 and 10($p < 0.05$). Additionally, ALOST gave the smallest mean MTRE ($p < 0.05$). It shows that ALOST has higher distinctiveness, which leads to a better registration performance. One may notice that in [34], [35], lower MTREs were found than the values reported by us. However, the referred methods rely on very different regularization techniques and are for this reason were not considered within the scope of our paper.

C. Registration of Abdominal MR Images

Ileocolonoscopy is considered the reference standard for the assessment of Crohn's Disease (CD). The disease activity is expressed in the Crohn's Disease Endoscopic Index of Severity, CDEIS [36]. This invasive procedure is considered very burdensome by many patients. Therefore, abdominal MR Imaging

is studied for diagnosing and grading CD. Rimola *et al.* [37] reported that the radiologic parameters to be evaluated include colon wall signal intensity and relative enhancement after contrast injection. The time intensity curves (TICs) obtained from dynamic contrast enhanced MRI (DCE-MRI) were also shown to contain important information on the inflammation of the bowel wall.

The quality of registration between the DCE-MRI and the structural post-contrast MRI data by each of the five methods (NMI, MIND_I, MIND_MP, MIND_PC, ALOST) was evaluated by: 1. calculating the average dice coefficient (DC) between ROIs from DCE-MRI and post-contrast MRI; 2. measuring the correlation between the measured image features after registration and the CDEIS score. We quantified the registration performance in an indirect way because no unique landmarks are identifiable in these data.

The abdominal MRI data employed in this paper were taken from a prior study into luminal Crohn's disease [38], which has been approved by the medical ethics committee. 30 out of 33 patients from the prior study gave written consent to use their data for future investigations. MR imaging included a free breathing $3D + time$ DCE series and a high-resolution post-contrast scan. Both were acquired by 3D T1-weighted spoiled gradient echo sequences performed in coronal planes. The DCE volumes contained $224 \times 224 \times 14$ voxels with a size of $1.78 \times 1.78 \times 2.5$ mm^3. In 6.1 minutes, 450 of such volumes were acquired at a rate of 0.8 second/volume. A contrast agent was injected after the 10th image volume was acquired. The image size of the post-contrast scans was $400 \times 400 \times 100$ voxels with a resolution of $1 \times 1 \times 2$ mm^3 and was acquired in a breath-hold. All patients underwent ileocolonoscopy within one month after the MRI scan was acquired. During ileocolonoscopy the CDEIS per bowel segment was scored [38].

Fig. 6 shows the image processing pipeline. The DCE-MRI volumes were first registered to themselves by another registration method to achieve correspondence over time [39], [40]. Therefore, an expiration template was created by a weighted average of expiration-phase volumes tentatively sampled from the whole series of DCE-MRI images. Then, the DCE-MRI volumes were retrospectively gated and registered to the template (see [40] for more information). After these steps, approximately 100 out of the 450 volumes remained.

Based on a routine radiological report, an expert annotated all ROIs (polygons around the areas affected by Crohn's disease)
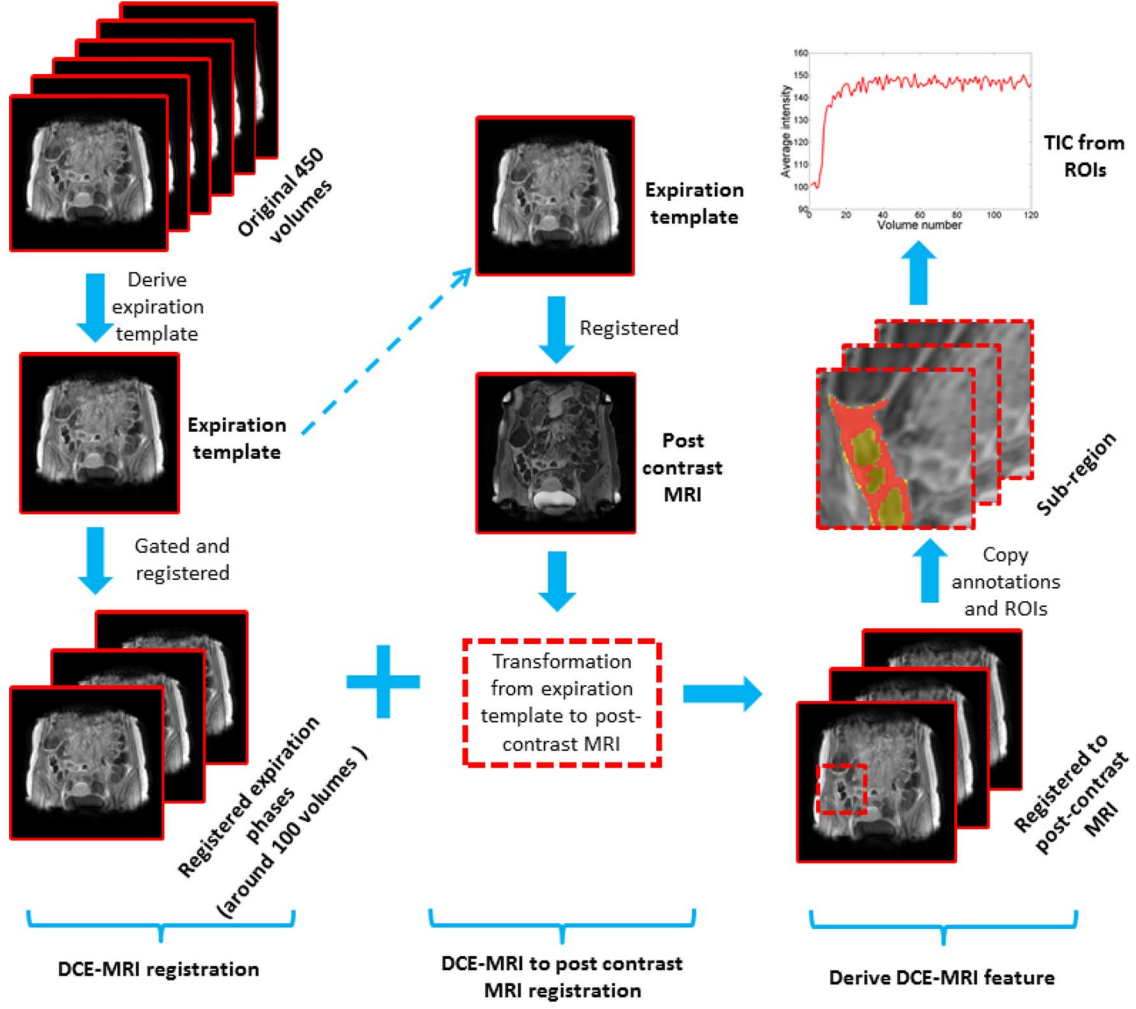
Fig. 6. Image processing pipeline for the registration of the DCE series to the post-contrast abdominal image. From left to right column: DCE-MRI gating and registration, DCE-MRI to post-contrast MRI registration, and extraction of DCE-MRI features (TIC, RCE) from a thresholded sub-region within the annotated ROI.

on the post-contrast MRI. Additionally, the affected bowel segment was identified: rectum (RT), left colon, i.e., sigmoid plus colon descendens (LC), transverse colon (TC), right colon, i.e., cecum plus colon ascendens (RC), and terminal ileum (TI).

Time intensity curves (TIC's) of the signal on the registered DCE series in the annotated ROIs (transferred from the post-contrast MRI after registration) contain information about the inflammation activity [39], [41] . Therefore, five methods were used to register the DCE expiration template to the post-contrast scan. Subsequently, the estimated transformation was applied to the entire gated and registered DCE series (Fig. 6, right column). Thereafter, the average signal over the ROI in each image from the DCE series might be determined. However, we noticed that the ROIs sometimes covered both bowel wall and lumen due to small registration errors. Since the bowel lumen has a low intensity compared to the bowel wall, the mean signal within the ROI was determined of only those pixels that exceeded a threshold. This threshold was determined automatically by means of the isodata thresholding algorithm [42]. Next we derived TICs from this ROI and fitted a bi-exponential model (based on Toft's model [43]):

$$s(t) = A_1 \exp(-\lambda_1 t) + A_2 \exp(-\lambda_2 t). \qquad (26)$$

The parameters are sorted such that A_1 reflects the strong upward slope (uptake of contrast) of the TIC and A_2 the slow downward slope (wash out of contrast). We choose parameter A_1 as an alternative to the relative contrast enhancement (RCE) between pre-contrast and post-contrast MRI which has been shown to be correlated to the disease severity [37].

1) A Visual Example of DCE to Post-Contrast Image Registration: The DCE-MRI and post-contrast MRI data differ both in resolution and in slab thickness (35 mm and 200 mm, respectively). Therefore, we first down-sampled the post-contrast MRI to $224 \times 224 \times 80$ voxels to get a similar voxel size as the DCE scans. We empirically found that the registration needed to be initialized close to the global optimum to avoid getting trapped in a local minimum. These local minima arise especially due to the large difference in slab thickness. To avoid this, we first rigidly registered the DCE-MRI expiration template to the post-contrast space using our proposed ALOST as the registration metric. After this initialization step, we applied NMI method and our four descriptors to register the (moving) expiration template to the (fixed) post-contrast image. A representative example is shown in Fig. 7. Observe that there are both large intensity differences and large spatial deformations between DCE-MRI and the post-contrast MRI.
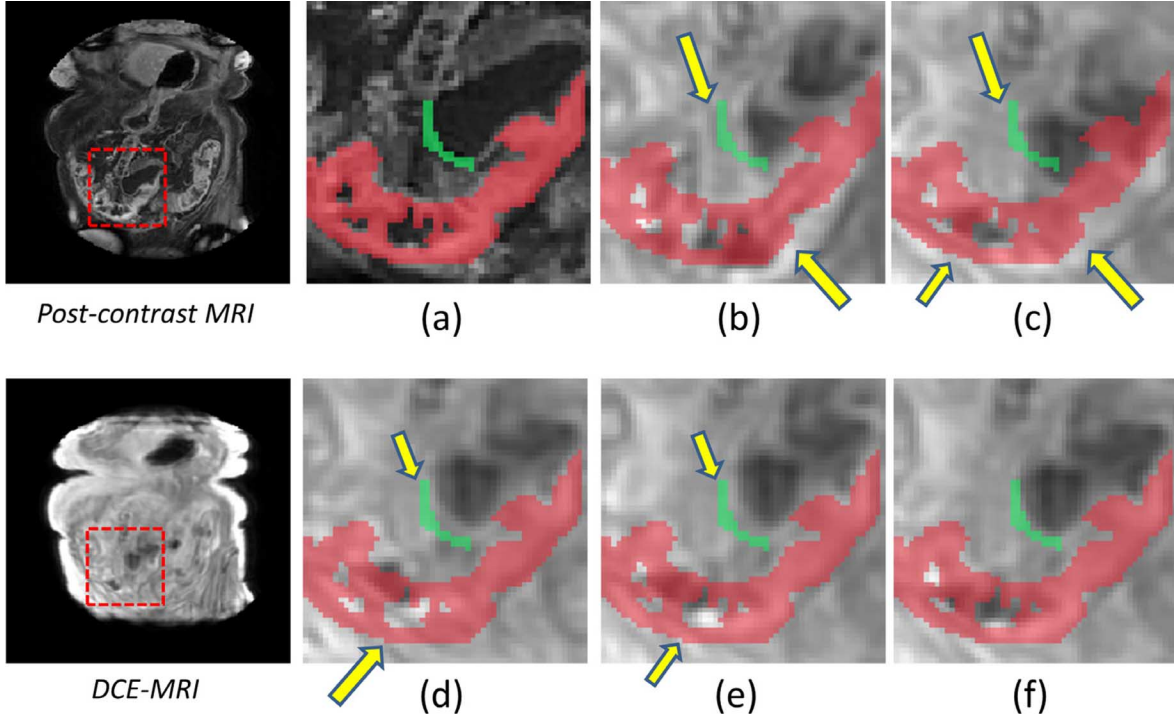
Fig. 7. An abdominal MR image showing a large bowel deformation in a post-contrast MR image (fixed image); DCE-MR image (moving image); (a) ROI in a sub-region of the post-contrast MR image (the red region is affected by Crohn's disease, the green region is healthy bowl wall); (b-f) same ROIs copied to the registration based on NMI (b), MIND_I (c), MIND_MP (d), MIND_PC (e), and ALOST (f). The yellow arrows point to a misregistration area in which the size of the arrow represents the amount of misregistration.
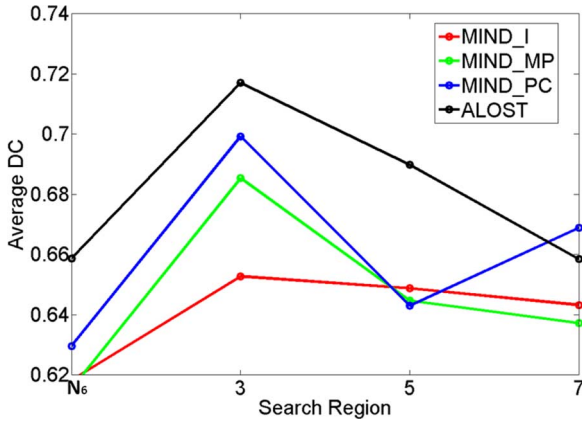

Fig. 8. Average Dice Coefficient (DC) of segmented regions in the DCE atlas and the post-contrast images as a function of the search region parameter R (R = N6 corresponds to a six-neighborhood region, R = 3 to a 3 × 3 × 3 region, R = 5 to a 5 × 5 × 5 region and R = 7 to a 7 × 7 × 7 region).
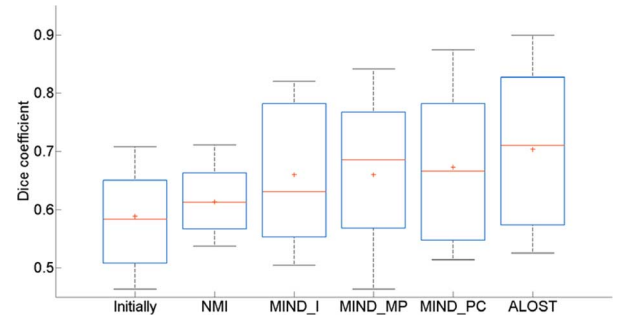

Fig. 9. Box plot of the average Dice Coefficient (DC) of segmented regions in the DCE atlas and the post-contrast images prior to registration (initially), and after registration based on the NMI, MIND_I, MIND_MP, MIND_PC and ALOST descriptors.
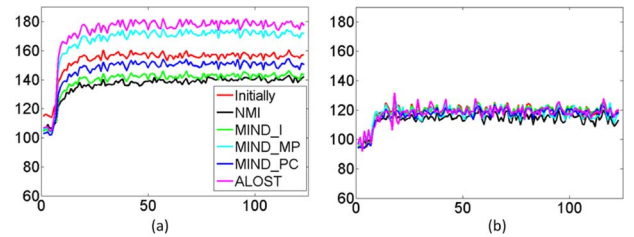

Fig. 10. TICs derived from the regions in Fig. 7 before registration (initially) and after registration base on the NMI, MIND_I, MIND_MP, MIND_PC and ALOST descriptors. (a): diseased region; (b): healthy region.

Fig. 7(a) shows segmented ROI affected by Crohn's disease in red as well as a healthy region in green. In Figs. 7(b)–7(f) show the annotations after registration by respectively NMI, MIND_I, MIND_MP, MIND_PC and ALOST.

NMI (Fig. 7(b)) and MIND_I (Fig. 7(c)) support only partial registration of the bowel boundary compared to MIND_MP and MIND_PC (Figs. 7(d) and 7(e)), pointed to by the arrow in the bottom-right corner. The size of the arrow represents the amount of misregistration. MIND_I fails to properly register this region as it is hampered by large intensity fluctuations and a large spatial deformation of the bowel. While MIND_MP and MIND_PC

improve here, their registration outcomes remain unsatisfactorily in the bottom left area of the image. We attribute this to indistinctiveness of features in soft tissue (MIND_MP) and the

TABLE III
CORRELATION COEFFICIENTS (CC) OF AN EXTRACTED DCE-MRI IMAGE FEATURE AND CDEIS (GROUND TRUTH) PRIOR TO REGISTRATION (INITIALLY), AND AFTER REGISTRATION BY NMI MIND_I, MIND_MP, MIND_PC AND ALOST. THE CONTRAST UPTAKE AMPLITUDE A1 WAS EXTRACTED FROM TIME INTENSITY CURVES THAT WERE CREATED BY EITHER ISODATA THRESHOLDING THE ROI IN EACH INDIVIDUAL DCE SCAN SEPARATELY (ROI-DCE) OR BY ISODATA THRESHOLDING ONLY THE POST-CONTRAST IMAGE (ROI-POST). P-VALUES OF THE CORRELATIONS ARE SHOWN BETWEEN BRACKETS. THE NUMBERS PRINTED IN BOLD FACE ARE THE MAXIMA (BEST RESULTS) OF EACH ROW

	Initially	NMI	MIND_I	MIND_MP	MIND_PC	ALOST
CC-post	0.30 (0.24)	0.39 (0.12)	0.41 (0.10)	0.49 (0.04)	0.44 (0.08)	**0.53 (0.03)**
CC-dce	0.39 (0.12)	0.41 (0.10)	0.43 (0.08)	0.52 (0.03)	0.50 (0.04)	**0.56 (0.02)**

presence of details (MIND_PC). Notice that the ALOST result (Fig. 7(f)) demonstrates a clear improvement compared to all other methods. Moreover, only ALOST gave good registration on both the healthy region (green) and the diseased region (red).

2) Evaluating Registration Accuracy by Dice Coefficient: It was found during colonoscopy that 6 out of 30 patients did not show signs of disease activity. In the remaining 24 patients 38 segments were identified showing marked disease activity: 16 TI, 9 RC, 3 TC, 5 LC, and 5 RT segments were given positive CDEIS scores. Only 17 out of the 38 segments were contained in the DCE-MRI scan *and* corresponded to the annotations in the post-contrast image: 14 TI, 2 TC, 1 LC segments. This reduction of segments resulted from (1) the limited slab size of the DCE scan, so that some diseased segments were not imaged; (2) differences in interpretation of the MR and colonoscopy images: several segments were identified to harbor Crohn's disease on one modality but not on the other, and vice versa.

We used two approaches to determine the ROIs from which TICs were generated: (1) ROI-dce: after duplicating the manual annotation to the DCE scans, we applied isodata thresholding to the annotated region for each DCE scan *separately*; (2) ROI-post: after applying isodata thresholding to the annotated region in the post-contrast image, we map the segmented region onto *all* DCE scans.

The search region $\mathbf{R}$ was first tuned to make sure that an optimal registration result was obtained for each method (as above). Therefore, we calculated the average dice coefficient (DC) between the ROI's-dce in the gated and registered DCE series and the ROI's-post for eight randomly chosen segments. We consider this a measure of the registration accuracy since a higher DC reflects a more similar ROI. The outcome is shown in Fig. 8, demonstrating that $\mathbf{R} = 3$ is the optimal search region for all methods. Observe that the ALOST method gave the largest average DC of all methods for almost each setting. Accordingly, we used $\mathbf{R} = 3$ for the methods on all 17 segments and calculated the average DC just as on the 'tuning' set. The outcome of this experiment is shown in Fig. 9. A two-sided signed rank test was applied to the average DC showing that ALOST had a significantly better outcome than NMI and MIND_I ($p < 0.05$). ALOST also outperforms the other methods, but these differences are not significant.

3) TICs From DCE- to Post-Contrast Image Registration: Fig. 10 shows the time intensity curves from the diseased (a) and healthy (b) regions from Fig. 7, before and after registration by the five methods. The TIC extracted from the diseased region after ALOST based registration gave the largest intensity change at the arrival of the contrast medium (Fig. 10(a)). The TICs extracted from the healthy region gave a similar but lower

transition. Since there is only little variation in the intensity of soft tissues in this region, a misregistration is not punished (Fig. 10(b)). Fig. 7(f) showed that the diseased bowel wall is nicely covered by the ROI after ALOST based registration. This is in contrast with the other approaches (cf. Figs. 7(b)–7(e)). Such a misregistration clearly affects the curves extracted from the diseased region. At the same time, there is an overlap of voxels between the ROIs after applying the different registrations. As a result, the fluctuations on the curves are highly correlated.

4) Correlating a Contrast Feature From DCE-MRI to CDEIS: Ultimately, it is our objective to predict the CDEIS from abdominal MRI data. The relative contrast enhancement, here represented by the A_1 parameter (see above), is considered an important parameter to do so [37]. Accordingly, we correlated A_1 to CDEIS as the final means to evaluate the NMI method and four descriptors. Since we introduced two types of ROI (ROI-dce and ROI-post), TICs were generated for both. Notice that in the former approach there is enhanced adaptation to local misregistration and/or intensity fluctuation by means of thresholding each DCE scan. The A_1 parameters were derived from both types of TICs separately. Table III shows the correlation coefficients (CC) from the five metrics generated as such. CC-dce represents the CC generated from ROI-dce and CC-post represents the CC generated from ROI-post.

Clearly, the ALOST metric combined with the CC-dce TIC generation gave the best correlation ($p < 0.05$). Unsurprisingly, all CC-dce based correlations are slightly higher than those based on CC-post since the former suffers less from misregistration. The number 0.56 is a reprehensive correlation value for a single feature as we used here. A similar correlation value can be found in [40]. It is well-known that for assessing Crohn's disease in MRI, more features should be measured such as the bowel wall thickness and the diseased segment length [37]. Similarly, colonoscopy scores are also based on multiple features. In future work, we aim to generate a new MRI score, which is based on multiple features.

IV. CONCLUSION

We developed a new descriptor called ALOST, which exploits the autocorrelation of the local image structure. The local structure is represented by a measure of the mean phase and the standard deviation (phase congruency) thereof. Both are derived from the monogenic signal representation. The ALOST descriptor has two important properties: (1) it is insensitive to intensity distortions; (2) it emphasizes 'salient' features in the image. These two properties make that image registration based on this descriptor improves compared to registration involving

other measures. This is particularly the case with data hampered by strong spatially varying intensity distortions and large spatial deformations.

We quantitatively compared the performance of a NMI-based method and four different descriptors on three different datasets. For the *synthetic abdominal MR images* 1) we calculated the average root mean square error (RMSE) of the displacement field on 80 image pairs after registration. The ALOST method gave the lowest average RMSE. From the *thoracic CT images* 2) we calculated the mean target registration error (MTRE) of 300 annotated landmarks on 10 datasets after registration. The ALOST method gave the smallest *mean* MTRE compared to the other descriptors. In the *abdominal MR images of patients with Crohn's disease* 3) we correlated an image feature representing the enhancement after contrast injection to an endoscopic measure of disease severity (CDEIS). It turned out that the registration by the ALOST method gave the highest correlation with CDEIS ($r = 0.56, p < 0.05$; the scatter plot could be found in supplementary figure). In all these experiments, ALOST outperformed the other methods. It shows that ALOST is a reliable technique especially to register data with space-variant intensity distortions. We expect that the method is also applicable to other images, such as retinal OCT images that suffer from luminance and contrast variations [29].

There are several limitations of our work. Our registration is based on a continuous optimization strategy. However, combining discrete optimization with a continuous one may give better registration by avoiding local minima [44]. Incorporating discrete optimization as proposed in [33] could also speed up the registration procedure. This will be required for clinical application. Sliding motion might also be a limitation for abdominal image registration, although we did not observe it in our images. If present, a potential solution might be to incorporate an explicit segmentation into our registration framework as in [44]. However, this would require accurate segmentation of the bowel, which is far from trivial. For instance, initial results by our group were reported in [39]. Another limitation of our work is in assessing the registration result of the abdominal MR data. The annotations of post-contrast MRI were drawn by only one expert. This might bias the results. Therefore, in future work, we are planning to include annotations from a second expert. The evaluation of our registration framework on a small number of ROIs is also a limitation of our work. We are currently testing our method on ROIs from a larger number of scans.

REFERENCES

[1] A. Sotiras, C. Davatazikos, and N. Paragios, "Deformable medical image registration: a survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of Mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[3] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Jul. 1997.

[4] D. Rueckert *et al.*, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[5] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recog.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.

[6] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[7] J. Pluim, J. Maintz, and M. Viergever, "Image registration by maximization of combined Mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000.

[8] H. Rivaz, Z. Karimaghaloo, and D. L. Collins, "Self-similarity weighted Mutual information: A new nonrigid image registration metric," *Med. Image Anal.*, vol. 18, no. 2, pp. 343–358, Feb. 2014.

[9] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 626–639, May 2006.

[10] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imag.*, vol. 29, 5, no. 1, pp. 19–29, May 2010.

[11] M. P. Heinrich *et al.*, "MIND: modality independent neighbourhood descriptor for multi-modal deformable registration," *Med. Image Anal.*, vol. 16, no. 7, pp. 1423–1435, Oct. 2012.

[12] D. De Nigris, D. L. Collins, and T. Arbel, "Multi-modal image registration based on gradient orientations of minimal uncertainty," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2343–2354, Dec. 2012.

[13] E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," *Proc. MICCAI*, pp. 726–733, 2006.

[14] M. P. Heinrich, M. Jenkinson, F. V. Gleeson, J. M. Brady, and J. A. Schnabel, "Deformable multimodal registration with gradient orientation based on structure tensors," *Ann. BMVA*, vol. 2011, no. 2, pp. 1–11, 2011.

[15] G. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Dordrecht, The Netherlands: Kluwer, 1995.

[16] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner, 1998.

[17] C. Wachinger and N. Navab, "Entropy and laplacian images: Structural representations for multi-modal registration," *Med. Image Anal.*, vol. 16, no. 1, pp. 1–17, Jan. 2012.

[18] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.

[19] M. Alessandrini, A. Basarab, H. Liebgott, and O. Bernard, "Myocardial motion estimation from medical images using the monogenic signal," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1084–1095, Nov. 2013.

[20] M. Mellor and J. M. Brady, "Phase mutual information as a similarity measure for registration," *Med. Image Anal.*, vol. 9, no. 4, pp. 330–43, Aug. 2005.

[21] A. Cifor, L. Risser, D. Chung, E. Anderson, and J. Schnabel, "Hybrid feature-based diffeomorphic registration for tumor tracking in 2-d liver ultrasound images," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1647–1656, Sep. 2013.

[22] A. Wong, D. Clausi, and P. Fieguth, "CPOL: Complex phase order likelihood as a similarity measure for MR-CT registration," *Med. Image Anal.*, vol. 14, no. 1, pp. 50–57, Feb. 2010.

[23] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," *Proc. CVPR*, pp. 60–65.

[24] M. P. Heinrich, M. Jenkinson, B. W. Papiez, J. M. Brady, and J. A. Schnabel, "Towards realtime multimodal fusion for image-guided interventions using self similarities," *Proc. MICCAI*, pp. 187–194, 2013.

[25] D. Boukerroui, J. A. Noble, and J. M. Brady, "On the choice of bandpass quadrature filters," *J. Math. Imag. Vis.*, vol. 21, no. 1-2, pp. 53–80, Jul. 2004.

[26] M. Morrone and R. Owens, "Feature detection from local energy," *Pattern Recogn. Lett.*, vol. 6, no. 5, pp. 303–313, Dec. 1987.

[27] P. Kovesi, "Edges are not just steps," *Proc. ACCV*, pp. 822–827, 2002.

[28] P. Kovesi, "Phase congruency detects corners and edges," *Proc. DICTA*, pp. 309–318, 2003.

[29] A. Myronenko and X. Song, "Image registration by minimization of residual complexity," *Proc. CVPR*, pp. 17–24, 2009.

[30] J. Midersitzki, *Numerical Methods for Image Registration*. Oxford, U.K.: Oxford Univ. Press, 2004.

[31] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[32] R. Castillo *et al.*, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.*, vol. 54, no. 7, pp. 1849–1870, Apr. 2009.

[33] H. P. Heinrich, M. Jenkinson, J. M. Brady, and J. A. Schnabel, "MRF-based deformable registration and ventilation estimation of lung CT," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1239–1248, Jul. 2013.

[34] L. Konig and J. Ruhaak, "A fast and accurate parallel algorithm for non-linear image registration using Normalized gradient fileds," *Proc. ISBI*, pp. 580–583, 2014.

[35] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," *Phys. Med. Biol.*, vol. 55, no. 1, pp. 305–327, Jan. 2010.

[36] J. Y. Mary and R. Modigliani, "Development and validation of an endoscopic index of the severity for Crohn's disease: A prospective multicentre study. Groupe d'Etudes thérapeutiques des affections inflammatoires du tube digestif (GETAID)," *Gut*, vol. 30, no. 7, pp. 983–989, Jul. 1989.

[37] J. Rimola *et al.*, "Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease," *Gut*, vol. 58, no. 8, pp. 1113–1120, Jan. 2009.

[38] M. L. Ziech *et al.*, "Dynamic contrast-enhanced MRI in patients with luminal Crohn's disease," *Eur. J. Radiol.*, vol. 81, no. 11, pp. 3019–3027, Nov. 2012.

[39] F. M. Vos *et al.*, "Computational modeling for assessment of IBD: to be or not to be?," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 3974–3977.

[40] Z. Li *et al.*, "Expiration-phase template-based motion correction of free-breathing abdominal dynamic contrast enhanced MRI," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, p. 1215, Apr. 2015.

[41] K. Horsthuis, C. Lavini, S. Bipat, P. C. Stokkers, and J. Stoker, "Perianal Crohn disease: Evaluation of dynamic contrast-enhanced MR imaging as an indicator of disease activity," *Radiology*, vol. 16, no. 10, pp. 1223–1230, May 2009.

[42] F. R. D. Velasco, "Thresholding using the ISODATA clustering algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 10, no. 11, pp. 771–774, Nov. 1980.

[43] P. S. Tofts, "Modeling tracer kinetics in dynamic GD-DTPA MR imaging," *J. Magn. Reson. Imag.*, vol. 7, no. 1, pp. 91–101, Jan. 1997.

[44] D. Kwon *et al.*, "PORTR: Pre-operative and post-recurrence brain tumor registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 651–667, Mar. 2014.