MIND: Modality Independent Neighbourhood Descriptor for Multi-Modal Deformable Registration

Mattias P. Heinrich[a,b,*], Mark Jenkinson[b], Manav Bhushan[a,b], Tahreema Matin[d], Fergus V. Gleeson[d], Sir Michael Brady[c], Julia A. Schnabel[a]

[a] *Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK*
[b] *Oxford University Centre for Functional MRI of the Brain, UK*
[c] *Department of Oncology, University of Oxford, UK*
[d] *Department of Radiology, Churchill Hospital, Oxford, UK*

Abstract

Deformable registration of images obtained from different modalities remains a challenging task in medical image analysis. This paper addresses this important problem and proposes a modality independent neighbourhood descriptor (MIND) for both linear and deformable multi-modal registration. Based on the similarity of small image patches within one image, it aims to extract the distinctive structure in a local neighbourhood, which is preserved across modalities. The descriptor is based on the concept of image self-similarity, which has been introduced for non-local means filtering for image denoising. It is able to distinguish between different types of features such as corners, edges and homogeneously textured regions. MIND is robust to the most considerable differences between modalities: non-functional intensity relations, image noise and non-uniform bias fields. The multi-dimensional descriptor can be efficiently computed in a dense fashion across the whole image and provides point-wise local similarity across modalities based on the absolute or squared difference between descriptors, making it applicable for a wide range of transformation models and optimisation algorithms. We use the sum of squared differences of the MIND representations of the images as a similarity metric within a symmetric non-parametric Gauss-Newton registration framework. In principle, MIND would be applicable to the registration of arbitrary modalities. In this work, we apply and validate it for the registration of clinical 3D thoracic CT scans between inhale and exhale as well as the alignment of 3D CT and MRI scans. Experimental results show the advantages of MIND over state-of-the-art techniques such as conditional mutual information and entropy images, with respect to clinically annotated landmark locations.

Keywords: Non-rigid registration, multi-modal similarity metric, self-similarity, non-local means, computed tomography, magnetic resonance imaging, pulmonary images

1. Introduction

The aim of medical image registration is to find the correct spatial mapping of corresponding anatomical or functional structures between images. Patient motion, due to different positioning or breathing level, and pathological changes between scans may cause non-rigid deformations, which need to be compensated for. Advances in recent years have resulted in a number of robust and accurate methods for deformable registration techniques for scans of the same modality, with registration accuracies close to the scan resolution (as demonstrated in an evaluation study of lung registration, Murphy et al. (2011)). However, the registration of images from different modalities remains a challenging and active area of research. Alignment of multi-modal images helps to relate clinically relevant and often complementary information from different scans. For example, it can be used in image guided interventions. Using multi-modal images can also help a clinician to make use of the complementary information present in different modalities and improve the diagnostic task. One common clinical application is the registration of computed tomography (CT) and magnetic resonance imaging (MRI), as it can combine the good spatial resolution and dense tissue contrast of a CT with the better soft tissue contrast of MRI.

In addition to the geometric distortion caused by patient motion, multi-modal registration also has to be able to deal with intensity distortions. Due to the different physical phenomena that are measured by the different modalities, there is no functional relation between the intensity mapping of corresponding anatomies. This problem can be addressed using geometric registration approaches, which aim to match a sparse set of descriptors, such as *scale invariant feature transform* (SIFT) (Lowe (1999)) or *gradient location and orientation histograms* (GLOH) (Mikolajczyk and Schmid (2005)), which are to some extent invariant to changes of intensity (or illumination) since they rely on image gradients and local orientations. However, they have not been successfully applied to multi-modal images, where the intensity variations are more severe. Voxel-wise in-

* Corresponding author:
Email address: mattias.heinrich@eng.ox.ac.uk (Mattias P. Heinrich)
URL: http://users.ox.ac.uk/~shil3388/ (Mattias P. Heinrich)

tensity based registration can also be used to align multi-modal images. This requires the use of a similarity metric derived from the image intensities that is robust to the non-functional intensity relationship.

Mutual information (MI), first introduced by Maes et al. (1997) and Viola and Wells III (1997), is an information theoretic measure, which aims to find a statistical intensity relationship across images and thereby maximises the amount of shared information between two images. For the rigid alignment of multi-modal images, MI has been very successful and is widely used (an overview is given in Pluim et al. (2003)). Its application to deformable multi-modal registration comes with many difficulties, and several weaknesses have been identified. The main disadvantage is that MI is intrinsically a global measure and therefore its local estimation is difficult, which can lead to many false local optima in non-rigid registration. Moreover, the optimisation of mutual information for non-rigid registration is computationally complex and converges slower than more simple intensity metrics, such as sum of squared differences (SSD), calculated over the intensities directly. Consequently, a new approach to deformable multi-modal registration has emerged, which uses a different scalar representation of both images based on a modality independent local quantity, such as local phase, gradient orientation or local entropy (Mellor and Brady (2005), Haber and Modersitzki (2006), Wachinger and Navab (2012)). These approaches benefit from their attractive properties for the optimisation of the cost function, since the point-wise (squared) differences can be used to minimise differences between the image representations. For challenging multi-modal scans it is however not always possible to find a scalar representation that is sufficiently discriminative.

In this article, we introduce a new concept for deformable multi-modal registration using a highly discriminative, multi-dimensional image descriptor, called the modality independent neighbourhood descriptor (MIND), which can be efficiently computed in a dense manner over the images and optimised using SSD. We make use of the concept of local self-similarity, which has been exploited in many different areas of image analysis, such as denoising (Buades et al. (2005)), super-resolution (Manjon et al. (2008)), image retrieval (Hörster and Lienhart (2008)), detection (Shechtman and Irani (2007)) and segmentation (Coupé et al. (2010)). It allows the formulation of an image descriptor, which is independent of the particular intensity distribution across two images and still provides a very good representation of the local shape of an image feature. It is based on the assumption that even though the intensity distribution of an anatomical structure may not correspond across modalities, it is reliable within a local neighbourhood in the same image. Therefore descriptors based on a simple intensity based metric, like SSD, can be extracted for each modality separately and then directly compared across images. The overview of our approach is schematically shown in Fig. 1. We first extract a dense set of high-dimensional image descriptors for both images independently based on the intensity differences within a search region around each voxel in the same modality. We embed this in a standard non-rigid registration framework to optimise the transformation parameters using a single-modal similarity metric (SSD), in order to compare descriptors across the two images.

This article extends our earlier work (Heinrich et al. (2011)) by using a more principled derivation of this image descriptor, thus making it more robust to changes in local noise and contrast and therefore allowing for the use of the L2 norm to compare descriptors across modalities. We also present a more thorough evaluation including quantitative comparisons to more recent multi-modal similarity metrics.

This paper is structured as follows: Section 2 presents an overview of related work in deformable multi-modal registration, as well as examples of the use of image self-similarity in literature. This includes a brief review of two recent techniques: conditional mutual information and entropy images, against which the proposed technique will be compared. Section 3 describes the formulation and implementation of MIND, demonstrating its sensitivity to different types of image features, such as corner points, edges and homogenous areas, and their local orientation. Details of its efficient implementation are presented, which greatly reduces the computational complexity by using convolution filters. The rigid and deformable registration framework used in the experiments, which is based on a multi-resolution Gauss-Newton optimisation, is presented in Section 4. Section 5 shows an evaluation of the robustness and accuracy of the presented method, first for the task of landmark detection in multi-modal 3D datasets under the influence of intensity distortions, then for deformable registration of CT lung scans, and finally on the clinical application of the alignment of volumetric CT and MRI scans of patients suffering from the lung disease empyema. The method's performance is quantitatively evaluated using gold standard landmarks localised by a clinical radiologist. Finally, the results are discussed and future research directions are given.

2. Background

2.1. Mutual information

Mutual information (MI) is derived from information theory and measures the statistical dependency of two random variables. It was first introduced to medical image registration for the rigid alignment of multi-modal scans by Maes et al. (1997) and Viola and Wells III (1997), and later used successfully in a variety of applications, including deformable registration (Rueckert et al. (1999), Meyer et al. (1997)). Studholme et al. (1999) introduced normalised mutual information (NMI) to cope with the effect of changing image overlap on MI. It is based on the assumption that a lower entropy of the joint intensity distribution corresponds to a better alignment.

An important disadvantage of mutual information for image registration is that it ignores the spatial neighbourhood of a particular voxel within one image and consequently, it does not use the spatial information shared across images. In the presence of image intensity distortions, such as a non-stationary bias field in an MRI scan, this can deteriorate the quality of the alignment, especially in the case of non-rigid registration

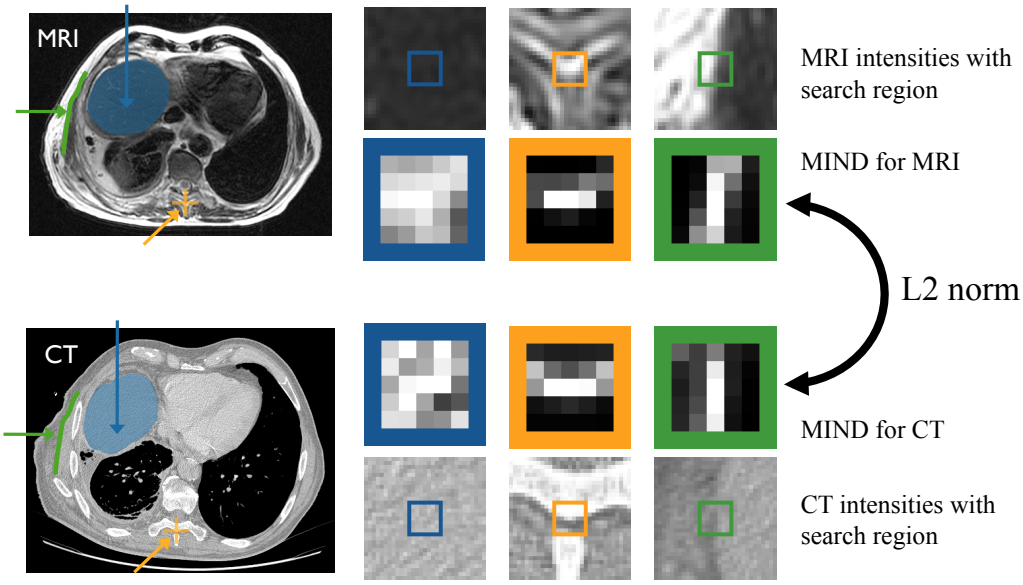


Figure 1: Proposed concept for the use of MIND for multimodal registration. MIND is calculated in a dense manner in CT and MRI. Three exemplary locations with different image features: ■ homogenous intensities (liver), ■ corner points at one vertebra and ■ image gradients at the boundary between fat and non-fat tissue are shown. The corresponding descriptors (in coloured boxes, high intensities correspond to small patch distances) are independent of the respective modality and can be easily compared using the L2 norm.

where the geometric constraints of the transformation are relaxed compared to rigid body alignment. One approach to overcome this problem is to include spatial information into the joint and marginal histogram computation. In Rueckert et al. (2000) a second-order mutual information measure is defined, which extends the joint entropy estimation to the spatial neighbours of a voxel and therefore uses a 4D histogram, where the third and forth dimensions define the probability of the spatial neighbours of a voxel to have a certain intensity. A problem that arises here is the curse of dimensionality, meaning that a lot of samples are needed to populate the higher-dimensional histogram. The authors therefore limit the number of intensity bins to 16, which might again decrease the accuracy. Three more recent approaches of MI including spatial context can be found in (Yi and Soatto (2011)), (Heinrich et al. (2012)) and (Zhuang et al. (2011)).

2.1.1. *Pointwise normalised mutual information*

In (Hermosillo et al. (2002), Rogelj et al. (2003)), variants of mutual information to obtain a pointwise similarity metric have been proposed. For the implementation of NMI as comparison method, the approach of (Rogelj et al. (2003)) is used in this work. The joint and marginal histograms p of two images I and J are obtained in a conventional manner by summing up the contribution of all intensity pairs to one global histogram. The local contribution $\mathrm{NMI}(\mathbf{x})$ for each voxel can then be obtained using:

$$\mathrm{NMI}(\mathbf{x}) = \log\left(\frac{p(I(\mathbf{x}),J(\mathbf{x}))}{p(I(\mathbf{x}))p(J(\mathbf{x}))}\right)\frac{1}{\sum_\Omega p(I(\mathbf{x}))\log p(I(\mathbf{x}))} \quad (1)$$

Alternatively, a local joint histogram estimation could be used, which however would limit the number of samples and would require more sophisticated histogram strategies like non-parametric windows (Dowson et al. (2008)), which are computationally extremely demanding for 3D volumes. A simplified computation for this technique was recently presented by Joshi et al. (2011).

2.1.2. *Conditional mutual information*

A number of disadvantages of using the traditional global MI approach have been analysed by Loeckx et al. (2010), Haber and Modersitzki (2006), and Studholme et al. (2006). These lie mainly in the sensitivity of MI (or NMI) to non-uniform bias fields in MRI. These can be often explained by the lack of spatial information in the joint histogram calculation. Different approaches have been proposed to include spatial context into MI as mentioned aboved. Studholme et al. (2006) introduce a third channel to the joint histogram containing a spatial or regional label. In this work, the recent approach called conditional mutual information (CMI), as introduced by Loeckx et al. (2010) is used for comparison purposes. In this technique, a third dimension is added to the joint histogram and a second dimension is added to the marginals representing the regional location of an intensity pair. The image is subdivided into a number of overlapping regions and each intensity pair only contributes to its specific regional histograms. A number of anchor points are evenly distributed on the image grid. Each voxel in a 3D volume is then attributed to its 8 nearest anchor points, and its contribution to this regional label $r(\mathbf{x})$ is weighted by the reciprocal spatial distance $w(I(\mathbf{x}),J(\mathbf{x}),\mathbf{x})$ to it. CMI is then defined as:

$$\mathrm{CMI}(\mathbf{x}) = -\sum_{\mathbf{x}\in\Omega} w(I(\mathbf{x}),J(\mathbf{x}),r(\mathbf{x}))\log\left(\frac{p(I(\mathbf{x}),J(\mathbf{x}))}{p(I(\mathbf{x}))p(J(\mathbf{x}))}\right) \quad (2)$$

In (Loeckx et al. (2010)) it was shown that this reduces the negative influence of bias fields and yields a higher registration accuracy for a small number of realistic test cases. The drawbacks lie again in the difficulty of populating this 3D histogram, and in the fact that corresponding anatomical structures, which are spatially further apart, are ignored.

2.2. Structural representation

A very different approach to multi-modal image registration is the use of a structural representation, which is assumed to be independent of a certain modality. One can then use a simple intensity-based measure across image representations. Using image gradients directly would be not representative across modalities, but the use of the local gradient orientation is possible and has been used in (Pluim et al. (2000)) for rigid registration and in (Haber and Modersitzki (2006), Heinrich et al. (2010)) and (De Nigris et al. (2010)) for deformable registration. In (Mellor and Brady (2005)), the local phase of the image was extracted using a technique called the monogenic signal, and further used for registration. However, in their work mutual information was used between phase images, which implies that there was still no direct dependency between the representations of different modalities. Our approach is different in that not a scalar representation, but a vector-valued image descriptor is derived for each voxel.

2.2.1. Entropy images

Local patch-based entropy images have been proposed by Wachinger and Navab (2012), which were then minimised using SSD across modalities, achieving similar registration accuracy as mutual information for rigid multimodal registration and some synthetic non-rigid experiments. The basic assumption that drives the registration based on entropy images is that intensity changes occur at the same locations in different modalities. An entropy image is produced by firstly calculating histograms of small image patches. The size p and weighting C_σ of the local patches is of great importance. The entropy value $E(\mathbf{x})$ for each voxel is then obtained using a Parzen Window smoothing of the histogram from which the Shannon entropy is calculated.

According to (Wachinger and Navab (2012)), the number of intensity bins for non-rigid registration should be sufficiently small to ensure a well populated local histogram, which however reduces the sensitivity to small intensity changes. A problem with this approach can be a changing level of noise within and across images - which in turn would influence the entropy calculation. The high complexity (p^d per voxel, where d is the dimension of the image) of the entropy image calculation could potentially be reduced using a convolution kernel for the contribution of each individual voxel to all neighbouring voxels within the size of a patch.

2.3. Self-similarity

Our approach uses the principle of self-similarity, a concept which has first been introduced in the domain of image denoising by Buades et al. (2005). These authors make use of similar image patches across a noisy image to obtain a noise-free pixel, which is computed as a weighted average of all other pixels in the image. The weights $w(i, j)$ used for the averaging are based on the sum of squared differences between the patch, which surrounds the pixel of interest, and all other patches in the image. The denoised pixels $NL(i, J)$ are then calculated using the following equation:

$$NL(i, J) = \sum_{j \in N} w(i, j) J(j) \tag{3}$$

where N is the neighbourhood of i. The approach demonstrated a very good performance for image denoising. The use of patches to measure similarity based on the weights $w(i, j)$ within the same image can easily capture a variety of image features, because it treats regions, edges, corners and textures in a unified way and is thus much more meaningful than using single intensities. In subsequent work, this approach was simplified to search for similar patches only within a smaller non-local search region (Coupé et al. (2006)). Figure 1 gives an example of how well the self-similarity pattern can describe the local structure around an image location. Mainly because of this property, the concept has been used later on in a variety of applications. Of particular interest is the application to object localisation by Shechtman and Irani (2007). Here, a correlation surface is extracted using colour patch distances and then stored in a log-polar histogram, which can be matched across images using the L1 norm.

3. Modality independent neighbourhood descriptor

In this section we will present the modality independent neighbourhood descriptor (MIND) and its use to define the similarity between two images based on the SSD of their descriptors. First we motivate the use of image self-similarity for the construction of an image descriptor. We will then propose the definition of self-similarity by using a Gaussian-weighted patch-distance and explain the spatial capture range of the descriptor.

3.1. Motivation and Concept

Our aim is to find an image descriptor, which is independent of the modality, contrast and noise level of images from different modalities and at the same time sensitive to different types of image features. Our approach is based on the assumption that a local representation of image structure, which can be estimated through the similarity of small image patches within one modality, is shared across modalities. As mentioned before, many different features may be used to derive a similarity cost function for image registration, such as corner points, edges, gradients, textures or intensity values. Figure 1 shows some examples on two slices of a CT and MRI volume.

Most intensity based similarity metrics employ only one of these features or need to define a specific combination of different features and a weighting between them. Image patches have been shown to be sensitive to very different types of image features including edges, points and texture. Using patches

for similarity calculations also removes the need for a feature specific weighting scheme. However, they are limited to single-modal images. In our approach, a multi-dimensional image descriptor, which represents the distinctive image structure in a local neighbourhood, is extracted based on patch distances for both modalities separately and afterwards compared using simple single-modal similarity measures.

MIND can be generally defined by a distance D_p, a variance estimate V and a spatial search region R:

$$\mathrm{MIND}(I, \mathbf{x}, \mathbf{r}) = \frac{1}{n} \exp\left(-\frac{D_p(I, \mathbf{x}, \mathbf{x}+\mathbf{r})}{V(I, \mathbf{x})}\right) \quad \mathbf{r} \in R \tag{4}$$

where n is a normalisation constant (so that the maximum value is 1) and $\mathbf{r} \in R$ defines the search region. By using MIND, an image will by represented by a vector of size $|R|$ at each location $\mathbf{x}$.

3.1.1. *Patch-based distance*

To evaluate Eq. 4 we need to define a distance measure between two voxels within the same image. As mentioned before, image patches offer attractive properties and are sensitive to the three main image features: points, gradients and uniformly textured regions. Therefore the straightforward choice of a distance measure $D_p(\mathbf{x}_1, \mathbf{x}_2)$ between two voxels $\mathbf{x}_1$ and $\mathbf{x}_2$ is the sum of squared differences (SSD) of all voxels between the two patches P of size $(2p+1)^d$ (with image dimension d) centred at $\mathbf{x}_1$ and $\mathbf{x}_2$.

$$D_p(I, \mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{p} \in P} (I(\mathbf{x}_1 + \mathbf{p}) - I(\mathbf{x}_2 + \mathbf{p}))^2 \tag{5}$$

The distance value defined in Eq. 5 has to be calculated for all voxels $\mathbf{x}$ in the image I and all search positions $\mathbf{r} \in R$. The naïve solution (which is e.g. used in Coupé et al. (2006)) would require $3(2p+1)^d$ operations per voxel and is therefore computationally very expensive.

We propose an alternative solution to calculate the exact patch-distance very efficiently using a convolution filter C of size $(2p+1)^d$. First a copy of the image I' is translated by $\mathbf{r}$ yielding $I'(\mathbf{r})$. Then the point-wise squared difference between I and $I'(\mathbf{r})$ is calculated. Finally, these intermediate values are convolved with the kernel C, which effectively substitutes the SSD summation in Eq. 5:

$$D_p(I, \mathbf{x}, \mathbf{x}+\mathbf{r}) = C \star (I - I'(\mathbf{r}))^2 \tag{6}$$

This procedure is now repeated for all search positions $\mathbf{r} \in R$. The solution of Eq. 6 is equivalent to the one obtained using Eq. 5. Using this method it is also easily possible to include a Gaussian weighting of the patches by using a Gaussian kernel C_σ of size $(2p+1)^d$. The computational complexity per patch distance calculation is therefore reduced from $(2p+1)^d$ to $d(2p+1)$ for an arbitrary separable kernel and $3d$ for a uniform patch weighting. A similar procedure has been proposed in the context of windowed SSD aggregation by Scharstein and Szeliski (1996).

3.2. *Variance measure for Gaussian function*

We want to obtain a high response for MIND for patches that are similar to the patch around the voxel of interest, and a low response for everything that is dissimilar. A Gaussian function (see Eq. 4) is used for this purpose. The denominator $V(I, \mathbf{x})$ in Eq. 4 is an estimation of the local variance. A smaller value for V yields a sharply decaying function, and higher values indicate a broader response. The parameter has to be related to the amount of noise in the image. The variance of the image noise can be estimated via pseudo-residuals ϵ calculated using a six-neighbourhood $\mathcal{N}$ (see Coupé et al. (2008)):

$$\epsilon_i = \sqrt{\frac{7}{6}}\left(I(\mathbf{x}_i) - \frac{1}{6}\sum_{\mathbf{x}_j \in \mathcal{N}} I(\mathbf{x}_j)\right) \tag{7}$$

ϵ is averaged over the whole image domain Ω to obtain a constant variance measure $V(I, \mathbf{x}) = \frac{1}{|\Omega|}\sum_{i \in \Omega} \epsilon_i^2$. This however increases the sensitivity of the image descriptors to spatially varying noise. Therefore a locally varying function would be beneficial. A better way of determining $V(I, \mathbf{x})$ is to use the mean of the patch distances themselves within a six-neighbourhood $\mathbf{n} \in \mathcal{N}$:

$$V(I, \mathbf{x}) = \frac{1}{6}\sum_{\mathbf{n} \in \mathcal{N}} D_p(I, \mathbf{x}, \mathbf{x}+\mathbf{n}) \tag{8}$$

Using this approach (Eq. 8), MIND can be automatically calculated without the need for any additional parameters.

Exemplary responses of the obtained descriptors for three different image features for both CT and MRI are shown in Fig. 1 (second and third row on the right), where a high intensity corresponds to a small patch distance. Fig. 1 demonstrates how well descriptors represent these features independent of modality.

3.3. *Spatial search region*

An important issue using MIND is the spatial extent of the search region (see R in Eq. 4) for which the descriptor is calculated. In the original work of Buades et al. (2005), self-similarity was defined across the whole image domain, thus coining the term: "non-local filtering". For the use in object detection, Shechtman and Irani (2007) used a sparse ensemble of self-similarity descriptors calculated with a search radius of 40 pixels, which was stored in a log-polar histogram. For the use of MIND in image registration, however, a smaller search region is sufficient. This can be explained by the prior knowledge of smooth deformations, which are enforced by the regularisation term of most deformable registration algorithms. We will define three different types of spatial sampling for the spatial search region R: dense sampling, sparse sampling (rays of 45 degrees), and a six-neighbourhood. Figure 2 illustrates these configurations, where the red voxel in the centre is the voxel of interest, and all gray voxels define R. The computational complexity is directly proportional to the number of sampled displacements, therefore the six-neighbourhood clearly offers the best time efficiency. If the neighbourhood is chosen too large, the resulting descriptor might be affected by non-rigid deformations.

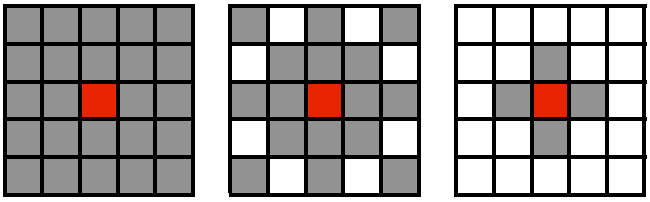


Figure 2: Different samplings of the search region: (a) dense, (b) sparse and (c) six-neighbourhood. Red voxel is the voxel of interest, gray voxels are being sampled $\mathbf{r} \in R$.

An evaluation of the influence of both patch-size (and weighting) and search region will be given in Section 5.2.1. A basic MATLAB implementation for the efficient calculation of MIND can be found in the electronic appendix.

3.4. Multi-modal similarity metric using MIND

One motivation for the use of MIND is that it allows to align multi-modal images using a simple similarity metric across modalities. Once the descriptors are extracted for both images, yielding a vector for each voxel, the similarity metric between two images is defined as the SSD between their corresponding descriptors. Therefore efficient optimisation algorithms, which converge rapidly can be used without further modification. We employ Gauss-Newton optimisation, which minimises the linearised error term in a least-square sense (Madsen et al. (1999)). In order to optimise the SSD of MIND, the similarity term $\mathcal{S}(\mathbf{x})$ of two images I and J at voxel $\mathbf{x}$ can be to be defined as the sum of absolute differences between descriptors:

$$\mathcal{S}(\mathbf{x}) = \frac{1}{|R|} \sum_{\mathbf{r} \in R} |\mathrm{MIND}(I, \mathbf{x}, \mathbf{r}) - \mathrm{MIND}(J, \mathbf{x}, \mathbf{r})| \qquad (9)$$

This requires $|R|$ computations to evaluate the similarity at one voxel. Some algorithms, especially discrete optimisation techniques (Glocker et al. (2008), Shekhovtsov et al. (2008)) use many cost function evaluations per voxel. In order to speed-up these computations the descriptor can be quantised to only 4 bit, without significant loss of accuracy. For $|R| = 6$ all possible distances between descriptors can be pre-computed and stored in a lookup-table.

The similarity $\mathcal{S}$ yields an intuitive display of the difference image after registration. Enabling single-modal similarity metrics by using an intermediate image representation is also the motivation in (Wachinger and Navab (2012)); in contrast to our work they reduce the alternative image representation to a single scalar value per voxel.

Our new similarity metric based on the MIND can be used in any registration algorithm with little need for further modification. We show in the experimental section that it can improve accuracy for both rigid and deformable registration of multi-modal data.

4. Gauss-Newton registration framework

This section describes the rigid and deformable registration framework, which will be used for all similarity metrics that are being compared in Section 5. We chose to use a Gauss-Newton optimisation scheme as it has an improved convergence compared to steepest descent methods (Zikic et al. (2010a)). For single-modal registration using SSD as similarity metric, Gauss-Newton optimisation is equivalent to the well known Horn-Schunck optical flow solution (Horn and Schunck (1981)) as shown in (Zikic et al. (2010b)).

4.1. Rigid registration

Rigid image registration aims to find the best transformation to align two images while constraining the deformation to be parameterised by a rigid-body (translation and rotation, 6 parameters). Extending this model to the more general affine transformation, the transformed location $\mathbf{x}' = (x', y', z')^T$ of a voxel $\mathbf{x} = (x, y, z)^T$ can be parameterised by $\mathbf{q} = (q_1, \ldots, q_{12})$:

$$\begin{aligned} u &= x' - x = q_1 x + q_2 y + q_3 z + q_{10} - x \\ v &= y' - y = q_4 x + q_5 y + q_6 z + q_{11} - y \\ w &= z' - z = q_7 x + q_8 y + q_9 z + q_{12} - z \end{aligned} \qquad (10)$$

where $\mathbf{u} = (u, v, w)^T$ is the displacement of $\mathbf{x}$. For a quadratic image similarity function $\mathbf{f}^2$, the Gauss-Newton method can be applied. It uses a linear approximation of the error term:

$$\begin{aligned} \mathbf{f}(\mathbf{x}') &\approx \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{u} \\ (\mathbf{J}^T\mathbf{J})\mathbf{u}_{\mathbf{gn}} &= -\mathbf{J}^T\mathbf{f} \end{aligned} \qquad (11)$$

where $\mathbf{J}(\mathbf{x})$ is the derivative of the error term with respect to the transformation and $\mathbf{u}_{\mathbf{gn}}$ is the update step. We insert Eqs. 10 into Eq. 11 and differentiate with respect to $\mathbf{q}$ to calculate $\mathbf{J}(\mathbf{x})$. The advantage of this method is that we can directly use the point-wise cost function derivatives with respect to $\mathbf{u}$ to obtain an affine transformation, so that MIND has to be computed only once per image.

Parameterizing a rigid-body transformation directly is more difficult. Therefore, at each iteration the best affine matrix is first estimated and then the best rigid-body transformation is found using the solution presented in Arun et al. (1987). The Gauss-Newton step is iteratively updated while transforming the source image towards the target. In order to speed up the convergence and avoid local minima, a multi-resolution scheme (with downsampling factors of 4 and 2) is used.

4.2. Diffusion-regularised deformable registration

Within the non-rigid registration framework, we aim to minimise the following cost function with respect to the deformation field $\mathbf{u} = (u, v, w)^T$, consisting of a non-linear similarity term $\mathcal{S}$ (dependent on $\mathbf{u}$) and a diffusion regularisation term:

$$\underset{\mathbf{u}}{\mathrm{argmin}} = \sum_{\mathbf{x}} \mathcal{S}\left(I_1(\mathbf{x}), I_2(\mathbf{x} + \mathbf{u})\right)^2 + \alpha\, \mathrm{tr}\left(\nabla\mathbf{u}(\mathbf{x})^T \nabla\mathbf{u}(\mathbf{x})\right) \qquad (12)$$

Since the objective function to be minimised is of the form $\sum_i \mathbf{f}_i^2$, we can again apply the Gauss-Newton optimisation method, where $\mathbf{f}$ is minimised iteratively with the update rule: $(\mathbf{J}^T\mathbf{J})\mathbf{u}_{\mathbf{gn}} = -\mathbf{J}^T\mathbf{f}$, where $\mathbf{J}$ is the derivative of $\mathbf{f}$ with respect to $\mathbf{u}$. This can be adapted to this regularised cost function. We simplify the notation to $\mathcal{S} = \mathcal{S}(I_1(\mathbf{x}), I_2(\mathbf{x}))$ and

$\nabla\mathcal{S} = (\frac{\delta\mathcal{S}}{\delta u}, \frac{\delta\mathcal{S}}{\delta v}, \frac{\delta\mathcal{S}}{\delta w})^T$ and $\Delta\mathbf{u} = \nabla(\nabla(\mathbf{u}(\mathbf{x}))$. The regularisation term is linear with respect to $\mathbf{u}$ as the differential operator is linear. The resulting update step given an initial or previous deformation field $\mathbf{u_{prev}}$ becomes then:

$$\left(\nabla\mathcal{S}^T\nabla\mathcal{S} - \alpha\Delta\right)\mathbf{u_{gn}} = -(\nabla\mathcal{S}^T\mathcal{S} - \alpha\Delta\mathbf{u_{prev}}) \quad (13)$$

Equation 13 is solved using successive over-relaxation (an iterative solver). The final deformation field is calculated by the addition of the update steps $\mathbf{u_{gn}}$. The parameter α balances the similarity term with the regulariser. The value of α has to be found empirically. This choice will be further discussed in Section 5.2.

4.3. Symmetric and inverse-consistent approach

For many deformable registration algorithms, there is a choice for one image to be the target and the other to be the source image. This places a bias on the registration outcome and may additionally introduce an inverse consistency error (ICE). The ICE has been defined by (Christensen and Johnson (2001)) for a forward transform A and a backward transform B to be the difference between AB^{-1} and the identity. In (Avants et al. (2008)) a symmetric deformable registration is presented, which calculates a transform from both images to a common intermediate image and also ensures that the forward transform is the inverse of the backward transform. The full forward transform transformation is calculated by $A(0.5) \circ B(0.5)^{-1}$, where 0.5 describes a transformation of half length (or with half the integration time, if velocity fields are used). We follow the same approach and estimate both A and B. We then use a fast iterative inversion method, as presented in (Chen et al. (2007)), to obtain $A(0.5)^{-1}$ and $B(0.5)^{-1}$. This approach helps to obtain diffeomorphic transformations, which means that no physically implausible folding of volume occurs. We use this symmetric approach in all deformable registration experiments.

5. Experiments

In this section we perform a number of challenging registration experiments to demonstrate the capabilities of MIND in medical image registration. We compare our new descriptor to state-of-the-art multi-modal similarity metrics: normalised mutual information (NMI), conditional mutual information (CMI), and SSD of entropy images (eSSD) within the same registration framework. We evaluate our findings based on the target registration error (TRE) of anatomical landmarks. The TRE for a given transformation $\mathbf{u}$ and an anatomical landmark pair $(\mathbf{x}, \mathbf{x}')$ is defined by (Maurer et al. (1997)):

$$\text{TRE} = \sqrt{(x + u(\mathbf{x}) - x')^2 + (y + v(\mathbf{x}) - y')^2 + (z + w(\mathbf{x}) - z')^2} \quad (14)$$

We first apply the different methods to landmark localisation within an aligned pair of T1 and PD weighted MRI scans of the Visible Human dataset. We then perform deformable registrations on ten CT datasets of lung cancer patients, and finally we register CT and MRI scans of patients with empyema.

5.1. Landmark localisation in visible human dataset

Evaluating multi-modal image registration in a controlled manner is not a trivial task. Finding and accurately marking corresponding anatomical landmarks across modalities is a difficult task even for a clinical expert. Random deformation experiments, as they are usually performed in the literature for multi-modal registration (e.g. in D'Agostino et al. (2003), Glocker et al. (2008), Mellor and Brady (2005), Wachinger and Navab (2012)), are mostly unrealistic. In order to perform a simulated deformation on multi-modal data, an aligned scan pair must be available, which is only usually possible for brain scans. Here the number of different tissue classes is a lot smaller than for chest scans, thus these experiments do not generalise very well. Moreover, simulated deformations hardly ever capture the complexity and physical realism of patient motion. To address these problems, we perform an alternative experiment: regional landmark localisation. For this purpose, we employ the less regularly used Visible Human dataset (VHD) (Ackerman (1998))[1]. Because the scans were taken post-mortem, no motion is present and different modalities are consequently in perfect alignment. We selected two MRI sequences, T1 and PD weighted volumes, as they offer a sufficient amount of cross-modality variations. The images are upsampled from their original resolution of 1.875x4x1.875 mm to form isotropic voxels of size 1.875 mm^3.

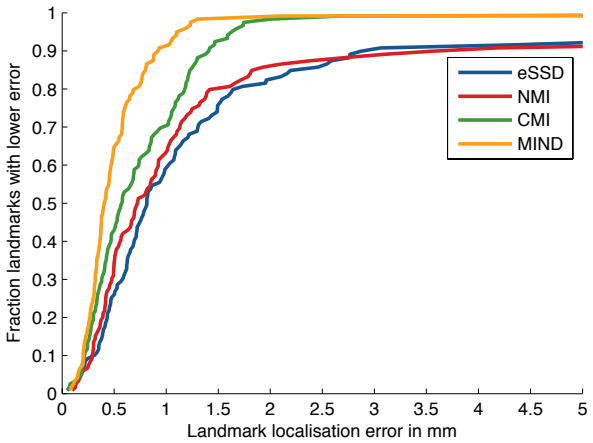


Figure 4: Cumulative distribution of landmark localisation error in mm for 119 landmarks located in the original T1/PD MRI scan of the Visible Human dataset. MIND achieves a significantly higher localisation accuracy.

In our tests we automatically select a large number (119) of geometric landmarks using the 3D version of the Harris corner detector (Rohr (2000)). Cross-sections of both sequences are shown in Fig. 3. For each landmark of the MRI-PD scan, we perform an exhaustive calculation of the similarity metric within a search window of 39x39x39 mm of the T1 image around the respective location. Since no regularisation is used in this experiment, we average the cost function over a local

[1]The Visible Human dataset is obtainable from `http://www.nlm.nih.gov/research/visible/getting_data.html`

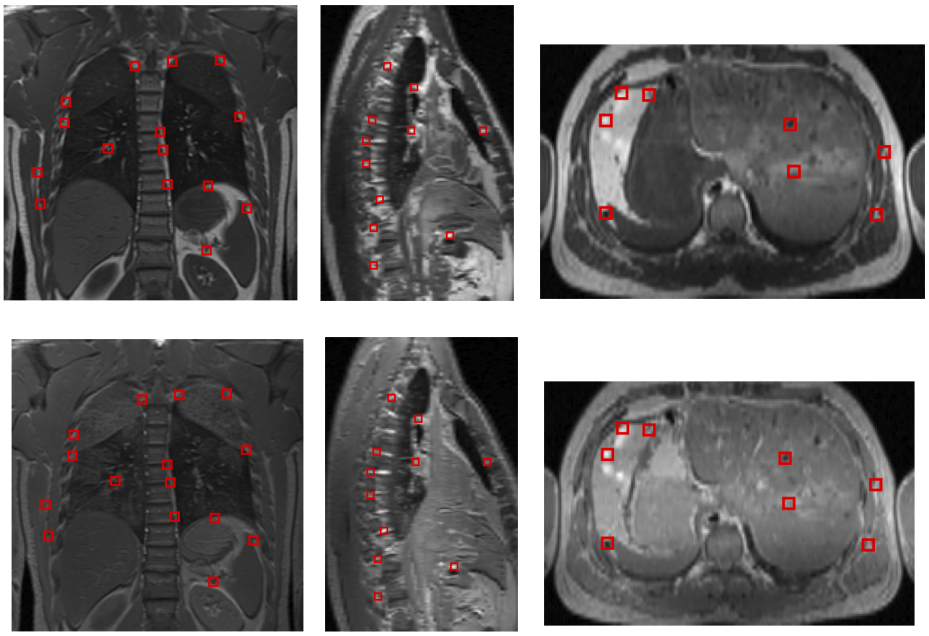


Figure 3: Visible Human Dataset used for landmark localisation experiment. T1 and PD MRI scan of post-mortem human are intrinsically aligned. The landmarks, which were used for evaluation are plotted with red squares.

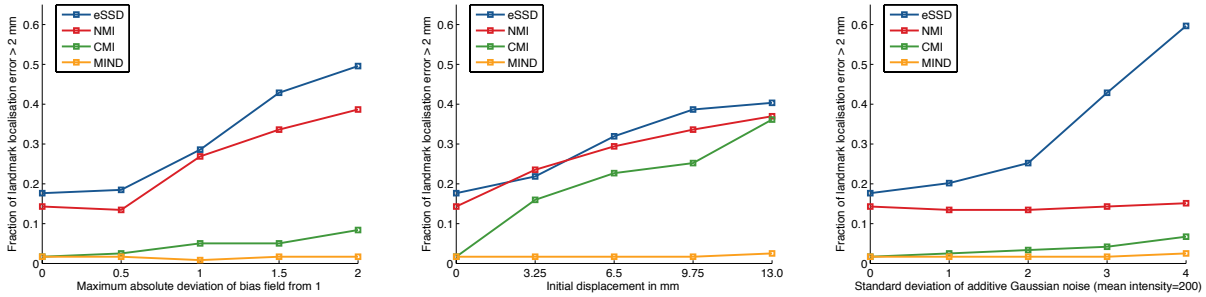


Figure 5: Fraction of falsely located landmarks (error > 2mm) for increasing bias field, initial misalignment (translation), and additive Gaussian noise in multi-modal pair of T1/PD MRI scan of Visible Human dataset. The resulting localisation deteriorates for NMI and eSSD with increased bias field. NMI, CMI and eSSD have a high localisation error for initially misaligned volumes. eSSD shows a high sensitivity to Gaussian noise.

neighbourhood with a radius 5 voxels. The optimal position (highest similarity) is calculated (up to subpixel accuracy) and compared to the known ground truth location. The Euclidean distance serves as localisation error. If the similarity metric is sufficiently discriminative, no other local optimum should appear within the search region. The distribution of the resulting error for all compared similarity metrics is shown in Fig. 4. MIND achieves a significantly lower localisation error than all other similarity metrics. We subsequently apply a non-uniform bias field (multiplicative linear gradient in y-direction), a translation, or additive Gaussian noise to the T1 scan. The fraction of falsely located landmarks with increasing image distortion is plotted in Fig. 5. MIND clearly outperforms both NMI and eSSD by achieving a consistently lower landmark localisation error. CMI is, as expected, not affected by the non-uniform bias field, however for an initial misalignment of the scan pair the joint histogram estimation becomes less reliable and the localisation accuracy deteriorates.

5.2. *Deformable registration of inhale and exhale CT scans*

We performed deformable registration on ten CT scan pairs between inhale and exhale phase of the breathing cycle, provided by the DIR-Lab at the University of Texas (Castillo et al. (2009))[2].The patients were treated for esophagus cancer, and a breathing cycle CT scan of thorax and upper abdomen was obtained, with slice thickness of 2.5 mm, and an in-plane resolution ranging from 0.97 to 1.16 mm. Even though this stipulates a single-modal registration problem, directly intensity based similarity criteria such as SSD may fail in some cases SSD may fail in some cases due to the changing appearance between inhale and exhale scans. Particular challenges for these registration tasks are the changing contrast between tissue and air, because the gas density changes due to compression (Castillo et al. (2010b)), discontinuous sliding motion between lung lobes and the lung rib cage interface, and large deformations of small features (lung vessels, airways). For each image 300 anatomical

[2]This dataset is freely available at `http://www.dir-lab.com`

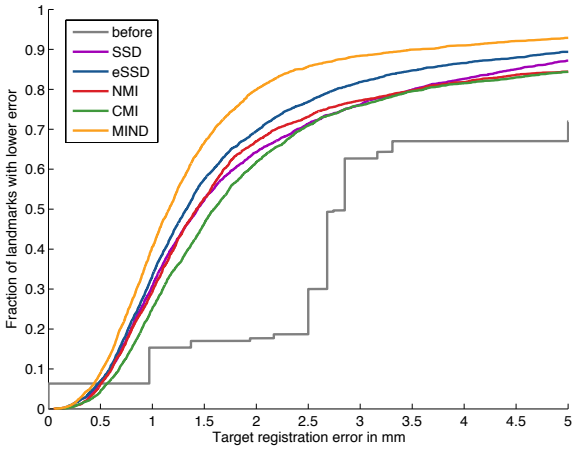


Figure 6: Deformable registration of 10 cases of CT scans evaluated with 300 expert landmarks per case. Registrations are performed between maximum inhale and exhale. The plot shows the cumulative distribution of target registration error, in mm. A significant improvement using MIND compared to all other methods has been found using a Wilcoxon rank sum test ($p<0.0001$). The staircase effect of TRE before registration is due to the voxel based landmark annotation.

landmarks have been carefully annotated by thoracic imaging experts with inter-observer errors of less than 1 mm. The maximum average landmark error before registration is 15 mm (for Case 8), the maximum displacement of a single landmark is 30 mm.

The cumulative distributions of target registration error (TRE) for all 3000 landmarks (all 300 landmarks for all 10 cases) after registration are shown in Figure 6. MIND achieves the lowest average and median TRE among all methods. The average error of the second best metric (eSSD) is more than a third higher. The Wilcoxon rank-sum test was used to compare the TRE between the different similarity metrics across all cases and for each case individually. We found a significant improvement for MIND compared to all other metrics. Entropy SSD could significantly improve the accuracy compared to NMI. A summary of the registration results is given in Table 1. The range of Jacobian values of the transformations are all positive, thus all deformation fields are free from singularities. An example of the registration outcome using our proposed method along with the magnitude of the deformation field is shown in Figure 7.

5.2.1. Choice of parameters

We used a symmetric three-level multiresolution scheme within the presented Gauss-Newton framework for all compared methods. The best parameters were carefully chosen based on the TRE obtained for Case 5. An overview is given in Table A.4 in the electronic appendix. The regularisation was chosen sufficiently high to ensure physically plausible transformations with no singularities (negative Jacobians). For CMI the spatial size of each regional label was set to be between 25^3 and 50^3 voxels, as suggested in (Loeckx et al. (2010)). The computation time for each 3D registration was between 4 and 5 min-

Table 1: Target registration error in mm for deformable registration of ten CT scans between inhale and exhale. Evaluation based on 300 manual landmark per case. Inter-observer error for landmark selection <1mm. A Wilcoxon rank test is performed between MIND and each comparison method. Cases, for which a significant improvement ($p<0.05$) was found are depicted below. As additional comparison, the results reported in the literature for two other techniques are shown below.

Metric		TRE (in mm)
before	mean±std	8.46±6.58
	quantiles [0.25, 0.5, 0.75]	[3.11, 6.97, 12.55]
	cases for which $p<0.05$	all
SSD	mean±std	2.73±3.72
	quantiles [0.25, 0.5, 0.75]	[0.89, 1.44, 2.85]
	cases for which $p<0.05$	all except 1,2
eSSD	mean±std	2.86±4.91
	quantiles [0.25, 0.5, 0.75]	[0.86, 1.33, 2.33]
	cases for which $p<0.05$	all except 1,2,4
NMI	mean±std	2.97±4.22
	quantiles [0.25, 0.5, 0.75]	[0.91, 1.42, 2.67]
	cases for which $p<0.05$	all except 2
CMI	mean±std	3.06±4.10
	quantiles [0.25, 0.5, 0.75]	[1.00, 1.59, 2.85]
	cases for which $p<0.05$	all
MIND	mean±std	2.14±3.71
	quantiles [0.25, 0.5, 0.75]	[0.77, 1.16, 1.79]
Results reported in literature		TRE (in mm)
Schmidt-Richberg et al. (2012)		
	direction-dependend	2.13±1.82
	diffusion regularisation	3.02±2.79
Castillo et al. (2010a)		1.35±1.43*

*These results are not directly comparable, as all frames of the 4D CT cycles are used during registration and more landmarks are evaluated.

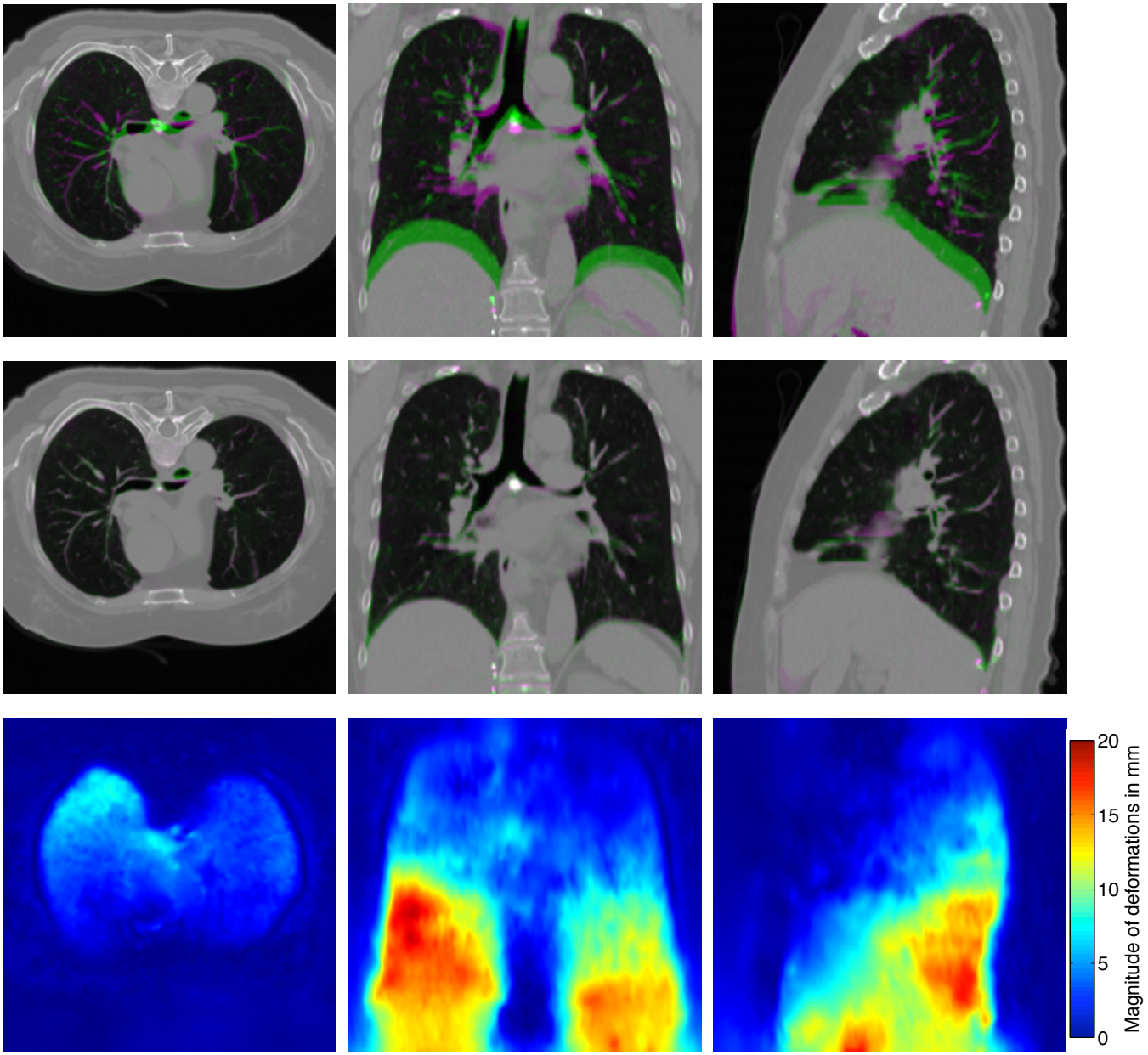


Figure 7: Deformable registration result for Case 5 of the CT dataset. Left: axial, middle: sagittal and right: coronal plane. Top row: before registration and centre row: after registration, using the proposed MIND technique. The target image is displayed in magenta and the source image in green (complementary colour). Bottom row shows the magnitude of the deformation field (red for large deformations) in mm.

utes for all methods (see Table 2). The influence of the choice of patch-size and search region for MIND has been evaluated using both single-modal and multi-modal registration tasks. Fig. 8 gives an overview of the obtained TRE. It can be generally seen that a Gaussian weighting $\sigma \approx 0.5$ (with a corresponding patch-size of 3x3x3) as well as a very small search region (six-neighbourhood yield a very high accuracy. For other applications with stronger image distortion and noise (e.g. ultrasound), we expect that larger patches and search regions would provide more robustness.

Table 2: Computation time (in seconds) for presented methods for Case 5 of CT dataset. For all metrics the SOR-solver for the Gauss-Newton optimisation takes 92 secs. The image dimensions are 256x256x106.

Metric	preprocessing	similarity term	full registration
	(for each GN iteration)		
eSSD	33.38	2.25	283.5
NMI	0.74	6.82	261.4
CMI	4.23	49.84	383.5
MIND	20.25	9.78	320.4

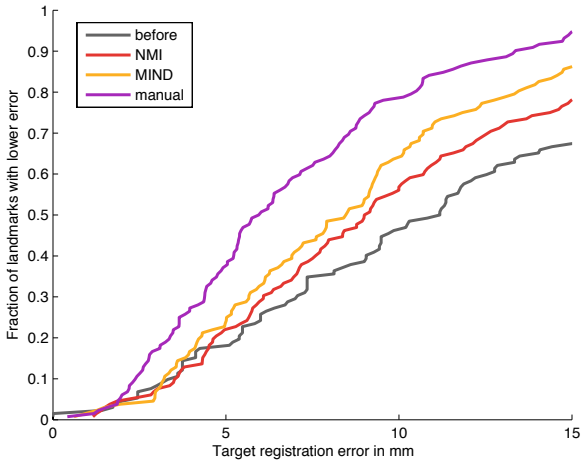


Figure 9: Rigid multi-modal registration of 11 cases of CT/MRI scans of empyema patients. Evaluated with 12 expert landmarks per case. The plot shows the cumulative distribution of target registration error, in mm. The manual registration error shows the residual error after a least square fit using a rigid transformation model to the ground truth landmark locations. MIND achieves an overall better performance than NMI.

5.3. Multi-modal registration of CT/MRI lung scans

Deformable multi-modal registration is important for a range of clinical applications. We applied our proposed technique to a clinical dataset of eleven patients, which were scanned with both CT and MRI. Different scanning protocols were employed for these clinical datasets. The CT volumes include scans with contrast, without contrast, and a CTPA (CT Pulmonary Angiogram) protocol. For the MRI scans, both T1-weighted and T2-weighted FSE-XL sequences within a single breath-hold were employed. All patients suffered from empyema, a lung disease characterised by infection of the pleura and excess fluid within the pleural space. The extra fluid may progress into an abscess and additionally, cause the adjacent lung to collapse and/or consolidate. Both modalities are useful for detecting this pathology, but because the patients are scanned in two different sessions and at different levels of breath-hold, there are non-rigid deformations, which makes it difficult for the clinician to relate the scans. The quality of the MRI scans is comparatively poor, due to motion artefacts, bias fields and a slice thickness of around 8 mm.

We asked a clinical expert to select manual landmarks for all eleven cases. 12 corresponding landmarks were selected in all image pairs, containing both normal anatomical locations and disease-specific places. It must be noted that some of the landmarks are very challenging to locate, both due to low scan quality and changes of the pathology in the diseased areas between scans. The intra-observer error has been measured to be 5.8 mm within the MRI and 3.0 mm within a CT scan.

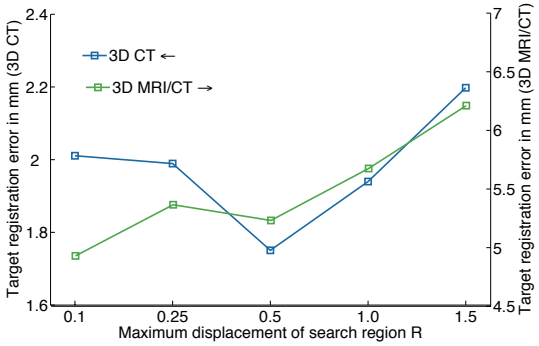
First a rigid registration of all cases using the proposed Gauss-Newton framework with the respective similarity metrics is performed. The resulting landmark errors are shown in Figure 9. MIND achieves a lower TRE of 9.3 mm, on average, compared to NMI (10.8 mm). We additionally calculated the optimal rigid body transformation using a least square fit of the ground truth landmark locations. We were not able to use entropy images for this multi-modal experiment as the structural representation is not sufficient to allow for the large variations in appearance and distortion between the CT and MRI scans and the registration fails for most cases (increased landmark error compared to ground truth after registration).

We use the rigid transformations obtained from the linear registration as initialisation of the subsequent deformable registration. For eSSD, the rigid transformations obtained using MIND, are employed as initialisation. The parameter choice for all compared methods can be found in Table A.5 in the electronic appendix.

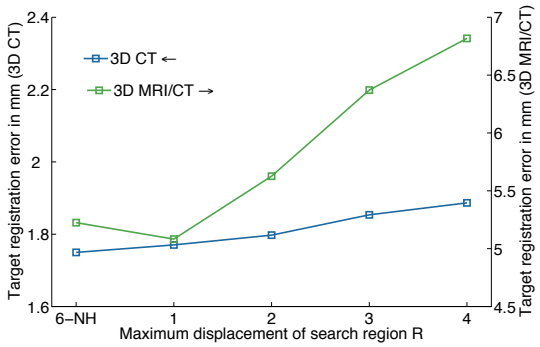
The obtained average TRE is 7.1 mm for MIND, 8.8 mm for CMI, 9.2 mm for NMI and 10.5 mm for eSSD. Even though the error for MIND is higher than what can be expected for a CT-to-CT registration, it is lower than the spatial resolution of the MRI scans and close to the intra-observer error. The distribution of landmark errors is shown in Fig. 10. Using a Wilcoxon rank test, a statistically significant improvement of MIND compared to NMI (p=0.019) and CMI (p=0.023) was found. An overview of the registration results is given in Table 3. The Jacobian values are all positive, thus no transformations contained any singularities. An example registration outcome for MIND and NMI is shown in Figure 11.

6. Discussion and Conclusion

We have presented a novel modality independent neighbourhood descriptor (MIND) for volumetric medical image registration. The descriptor can be efficiently computed locally across the whole image, and it allows for accurate and reliable alignment in a variety of registration tasks. Compared to mutual information it does not rely on the assumption of a global (or re-



(a) Increasing Gaussian weighting σ for patch-distance (see Sec. 3.1.1), the half-size of the patch is $p = \lceil 1.5\sigma \rceil$



(b) The maximum displacement $\mathbf{r}_{\max}$ of the search region R (see Sec. 3.3). 6-NH stands for a six-neighbourhood.

Figure 8: Parameter variation for MIND to determine the best choice of (a) σ in patch-distance D_p and (b) the spatial search region R. The TRE is evaluated for one single-modal (3D CT) case (left y-axis) and one multi-modal (3D MRI/CT) registration (right y-axis). Based on these tests, we choose $\sigma = 0.5$ and a six-neighbourhood for all experiments.

Table 3: Target registration error in mm for deformable registration of eleven CT/MRI scan pairs of empyema patients. Evaluation based on 12 manual landmarks per case. Slice thickness of MRI scans is 8 mm (in-plane resolution $\approx$ 1mm), intra-observer error for landmark localisation is 5.8 mm in MRI scans. A Wilcoxon rank test has been performed between presented methods. Significant improvements using MIND are found compared any other method (using all 132 landmarks).

Metric		TRE (in mm)
before	mean±std	**13.49±10.53**
	quantiles [0.25, 0.5, 0.75]	[6.00, 11.18, 17.63]
	p-value	$< 10^{-6}$
eSSD	mean±std	**10.49±6.78**
	quantiles [0.25, 0.5, 0.75]	[5.43, 9.34, 13.79]
	p-value	$< 10^{-5}$
NMI	mean±std	**9.18±7.40**
	quantiles [0.25, 0.5, 0.75]	[4.06, 6.91, 11.84]
	p-value	< 0.019
CMI	mean±std	**8.79±6.51**
	quantiles [0.25, 0.5, 0.75]	[3.84, 7.01, 11.87]
	p-value	< 0.023
MIND	mean±std	**7.12±5.88**
	quantiles [0.25, 0.5, 0.75]	[3.33, 5.68, 9.10]

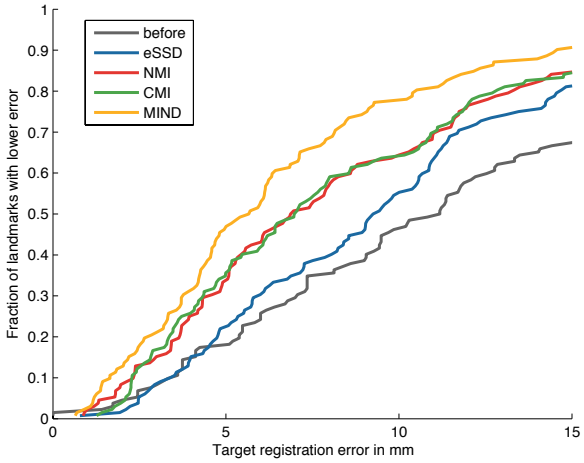
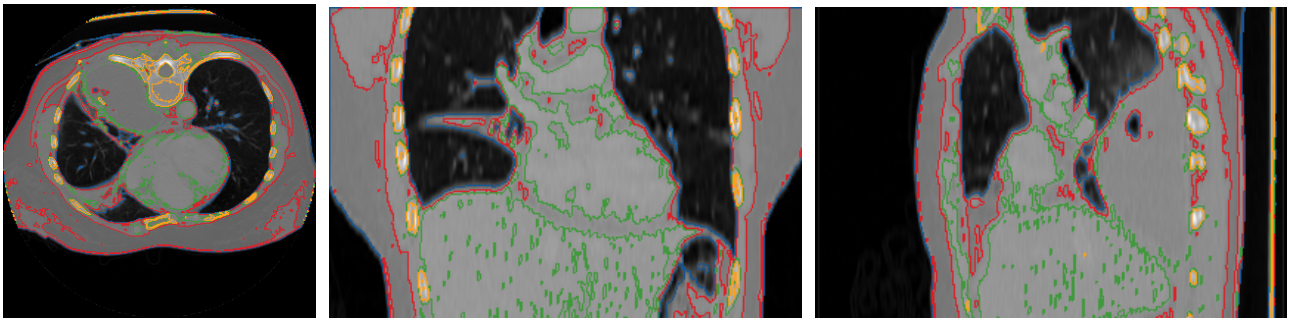
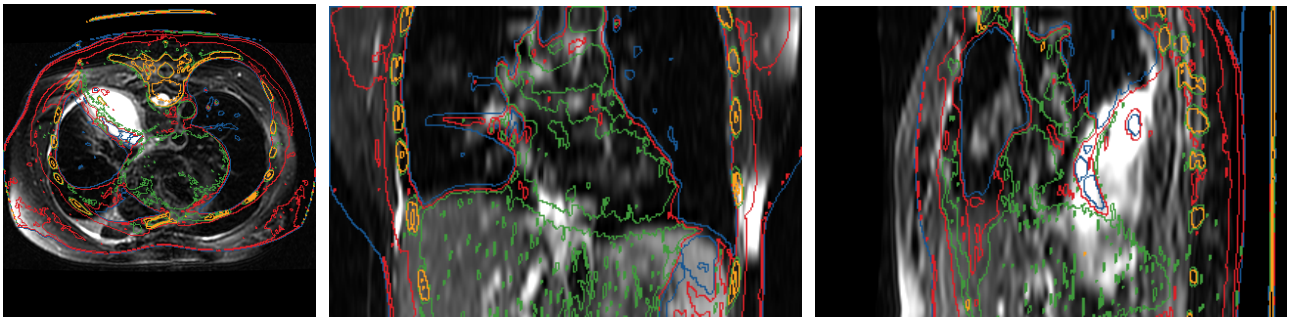


Figure 10: Deformable multi-modal registration of 11 cases of CT/MRI scans of empyema patients, evaluated with 12 expert landmarks per case. The plot shows the cumulative distribution of target registration error, in mm. MIND achieves a statistically significant (p<0.023) better result than all other methods. The comparatively high residual error is due to both low scan quality, (in-plane resolution is $\approx$ 1mm, but the slice thickness up to 8 mm) and the challenging landmark selection for the clinical expert (intraobserver error is 5.8 mm).

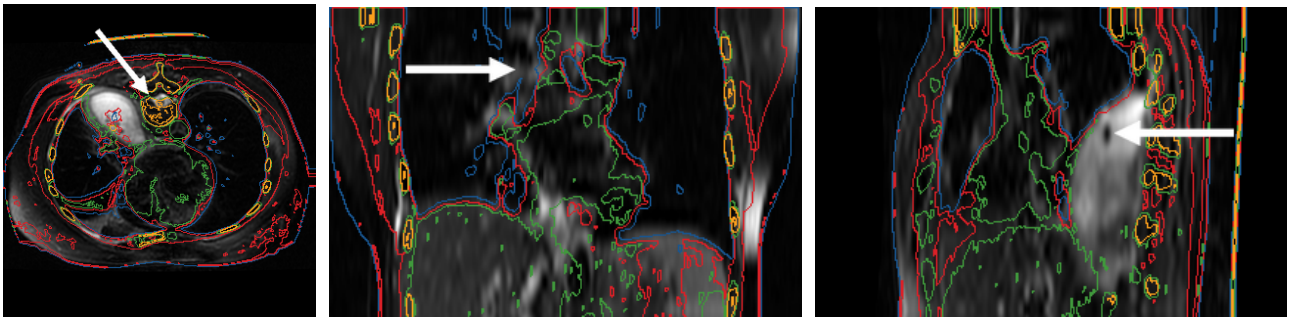
gional) intensity relation. The negative influence of initial misalignment and non-uniform bias fields is massively reduced and the difficult task of setting the correct parameters for the histogram calculation can be avoided. Apart from the regularisation parameter, a standard setting can be used for all registration tasks. The descriptor is not rotationally invariant, which might be a limitation in the case of strong rotations. However, the sensitivity of MIND to the local orientation may in fact lead to improved accuracy as suggested by the previous work of Pluim et al. (2000) and Haber and Modersitzki (2006). The modality independent representation using a vector based on the local neighbourhood (which allows it to capture orientation) instead of a scalar value (used in entropy images) shows clear improvements for real multi-modal registration experiments. The implementation is straightforward, the running time comparable



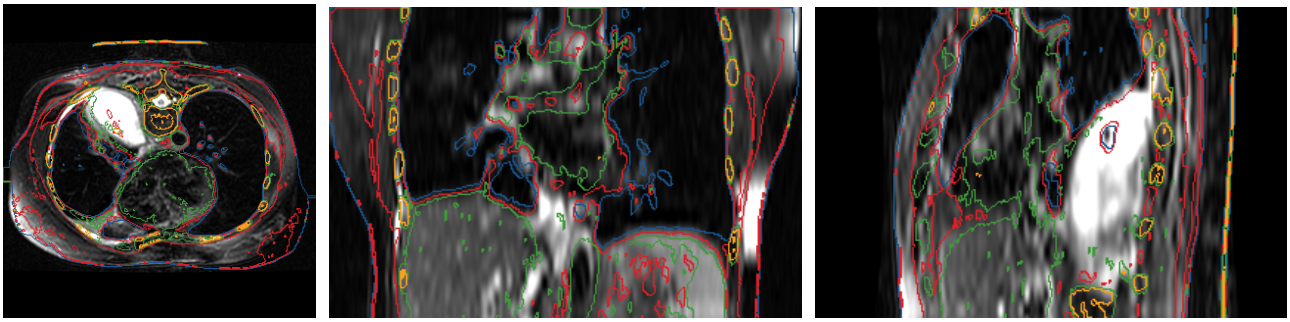
(a) CT scan of empyema patient with 4 relevant contour plots to guide the visualisation of registration results.



(b) MRI scan with identical CT contour plots before registration.



(c) Identical MRI scan with CT contour plots deformed according to non-rigid registration using NMI. The white arrows depict inaccurate registration close to one vertebrae, the inner lung boundary and gas pocket in empyema.



(d) Identical MRI scan with CT contour plots deformed according to non-rigid registration using MIND. A visually better alignment could be achieved.

Figure 11: Deformable CT/MRI registration results for Case 11 of empyema dataset. Left: axial, middle: sagittal and right: coronal plane. The third row shows the registration outcome using NMI. A better alignment is obtained when using MIND (forth row).

to other methods, and an important advantage of MIND is that it is calculated point-wise and can therefore be adapted to almost any registration algorithm.

We performed an extensive evaluation of our proposed method and three state-of-the-art multi-modal similarity metrics: entropy images, normalised and conditional mutual information. Tables 1 and 3 summarise the deformable registration results on two very challenging datasets. The results clearly demonstrate the advantages of the proposed descriptor. MIND achieves a higher accuracy and more robust correspondences for the CT dataset. The application of deformable registration to multi-modal medical images has so far remained a less sophisticated and advanced field with very few published results on clinically relevant data. Our proposed descriptor marks a novel contribution to this area. We verified its robustness to noise, field inhomogeneities and complex intensity relations in two experiments. First, the localisation of geometric landmarks was tested in an intrinsically aligned T1/PD MRI scan pair of the Visible Human dataset. Here the high discrimination and independency of bias fields has been demonstrated. Secondly for the deformable registration of clinical CT and MRI scans, we found a significant improvement over all other tested metrics.

While our validation was focused on CT and MRI modalities, we believe that our approach generalises well and further use could be made of this concept in a variety of medical image registration tasks. The application of MIND to other multi-modal registration tasks, such as registration of PET, contrast enhanced MRI and ultrasound, also to other anatomical regions, is subject for future work. A limitation of our approach is that it requires an anatomical feature to be present in both modalities, if this assumption is violated the concept of mutual-saliency (Ou et al. (2011)) could be incorporated to improve the robustness in these cases.

Further improvements might be possible. The use of more sophisticated deformation models could address application-specific challenges, such as slipping organ motion (Schmidt-Richberg et al. (2012)) and bladder filling or bowel gases (Foskey et al. (2005)). Employing a different optimisation scheme such as a registration based on MRF labelling (Glocker et al. (2011)), may allow us to find better maxima in the similarity function. In future work we will investigate the potential advantages of incorporating MIND into a discrete optimisation technique.

Acknowledgments

The authors would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre. J.A.S. acknowledges funding from EPSRC EP/H050892/1.

References

Ackerman, M., 1998. The visible human project. Proceedings of the IEEE 86, 504 –511.

Arun, K.S., Huang, T.S., Blostein, S.D., 1987. Least-squares fitting of two 3-d point sets. Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-9, 698 –700.

Avants, B., Epstein, C., Grossman, M., Gee, J., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26 – 41.

Buades, A., Coll, B., Morel, J.M., 2005. A non-local algorithm for image denoising, in: CVPR 2005, IEEE Computer Society. pp. 60–65.

Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T., 2010a. Four-dimensional deformable image registration using trajectory modeling. Physics in Medicine and Biology 55, 305.

Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T., 2009. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine and Biology 54, 1849.

Castillo, R., Castillo, E., Martinez, J., Guerrero, T., 2010b. Ventilation from four-dimensional computed tomography: density versus jacobian methods. Physics in Medicine and Biology 55, 4661.

Chen, M., Lu, W., Chen, Q., Ruchala, K.J., Olivera, G.H., 2007. A simple fixed-point approach to invert a deformation field. Medical Physics 35, 81.

Christensen, G., Johnson, H., 2001. Consistent image registration. IEEE Transactions on Medical Imaging 20, 568 –582.

Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D., 2010. Nonlocal patch-based label fusion for hippocampus segmentation, in: Jiang, T., Navab, N., Pluim, J., Viergever, M. (Eds.), Medical Image Computing and Computer-Assisted Intervention MICCAI 2010. Springer Berlin / Heidelberg. volume 6363 of *Lecture Notes in Computer Science*, pp. 129–136.

Coupé, P., Yger, P., Barillot, C., 2006. Fast non local means denoising for 3D MR images. MICCAI 2006 , 33–40.

Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. Medical Imaging, IEEE Transactions on 27, 425 –441.

D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P., 2003. A viscous fluid model for multimodal non-rigid image registration using mutual information. Medical Image Analysis 7, 565–575.

De Nigris, D., Mercier, L., Del Maestro, R., Collins, D.L., Arbel, T., 2010. Hierarchical multimodal image registration based on adaptive local mutual information, pp. 643–651.

Dowson, N., Kadir, T., Bowden, R., 2008. Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, 1841 –1857.

Foskey, M., Davis, B., Goyal, L., Chang, S., Chaney, E., Strehl, N., Tomei, S., Rosenman, J., Joshi, S., 2005. Large deformation three-dimensional image registration in image-guided radiation therapy. Physics in Medicine and Biology 50, 5869.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008. Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12, 731 – 741.

Glocker, B., Sotiras, A., Komodakis, N., Paragios, N., 2011. Deformable medical image registration: Setting the state of the art with discrete methods. Annual Review of Biomedical Engineering 13, 219–244.

Haber, E., Modersitzki, J., 2006. Intensity gradient based registration and fusion of multi-modal images. MICCAI 2006 , 726–733.

Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, J., Schnabel, J., 2011. Non-local shape descriptor: A new similarity metric for deformable multi-modal registration, in: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing and Computer-Assisted Intervention MICCAI 2011. Springer Berlin / Heidelberg. volume 6892 of *Lecture Notes in Computer Science*, pp. 541–548.

Heinrich, M., Jenkinson, M., Brady, M., Schnabel, J., 2012. Textural mutual information based on cluster trees for multimodal deformable registration, in: Biomedical Imaging: From Nano to Macro, 2012 IEEE International Symposium on, pp. 1–4.

Heinrich, M., Schnabel, J., Gleeson, F., Brady, M., Jenkinson, M., 2010. Non-Rigid Multimodal Medical Image Registration using Optical Flow and Gradient Orientation. Proc. Medical Image Analysis and Understanding , 141–145.

Hermosillo, G., Chefd'hotel, C., Faugeras, O., 2002. Variational methods for multimodal image matching. Int. J. Comput. Vision 50, 329–343.

Horn, B., Schunck, B., 1981. Determining optical flow. Artificial Intelligence 17, 185–203.

Hörster, E., Lienhart, R., 2008. Deep networks for image retrieval on large-scale databases, in: Proceeding of the 16th ACM international conference

on Multimedia, ACM, New York, NY, USA. pp. 643–646.

Joshi, N., Kadir, T., Brady, S., 2011. Simplified computation for nonparametric windows method of probability density function estimation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33, 1673 –1680.

Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P., 2010. Nonrigid image registration using conditional mutual information. Medical Imaging, IEEE Transactions on 29, 19 –29.

Lowe, D., 1999. Object recognition from local scale-invariant features, in: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, pp. 1150 –1157 vol.2.

Madsen, K., Bruun, H., Tingleff, O., 1999. Methods for non-linear least squares problems.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 16, 187–198.

Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., Garca-Mart, G., Mart, L., 2008. MRI denoising using non-local means. Medical Image Analysis 12, 514 – 523.

Maurer, C.R., J., Fitzpatrick, J., Wang, M., Galloway, R.L., J., Maciunas, R., Allen, G., 1997. Registration of head volume images using implantable fiducial markers. Medical Imaging, IEEE Transactions on 16, 447 –462.

Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for registration. Medical Image Analysis 9, 330 – 343.

Meyer, C.R., Boes, J.L., Kim, B., Bland, P.H., Zasadny, K.R., Kison, P.V., Koral, K., Frey, K.A., Wahl, R.L., 1997. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. Medical Image Analysis 1, 195 – 206.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, 1615 –1630.

Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M., Schnabel, J., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J., Ourselin, S., Muenzing, S., Viergever, M., De Nigris, D., Collins, D., Arbel, T., Peroni, M., Li, R., Sharp, G., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J., Staring, M., Klein, S., Stoel, B., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D., Pluim, J., 2011. Evaluation of registration methods on thoracic ct: The empire10 challenge. Medical Imaging, IEEE Transactions on 30, 1901 –1920.

Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., 2011. Dramms: Deformable registration via attribute matching and mutual-saliency weighting. Medical Image Analysis 15, 622 – 639. Special section on IPMI 2009.

Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registration of medical images: a survey. Medical Imaging, IEEE Transactions on 22, 986 –1004.

Pluim, J., Maintz, J.B., Viergever, M., 2000. Image registration by maximization of combined mutual information and gradient information. MICCAI 2000 , 103–129.

Rogelj, P., Kovacic, S., Gee, J.C., 2003. Point similarity measures for non-rigid registration of multi-modal data. Comput. Vis. Image Und. 92, 112 – 140.

Rohr, K., 2000. Elastic registration of multimodal medical images: A survey. Künstliche Intelligenz 14, 11–17.

Rueckert, D., Clarkson, M.J., Hill, D.L.G., Hawkes, D.J., 2000. Non-rigid registration using higher-order mutual information, SPIE. pp. 438–447.

Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. Medical Imaging, IEEE Transactions on 18, 712 –721.

Scharstein, D., Szeliski, R., 1996. Stereo matching with non-linear diffusion, in: Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on, pp. 343 –350.

Schmidt-Richberg, A., Werner, R., Handels, H., J., E., 2012. Estimation of slipping organ motion by registration with direction-dependent regularization. Medical Image Analysis 16, 150–159.

Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos, in: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pp. 1 –8.

Shekhovtsov, A., Kovtun, I., Hlavac, V., 2008. Efficient MRF deformation model for non-rigid image matching. Computer Vision and Image Understanding 112, 91 – 99. Special Issue on Discrete Optimization in Computer Vision.

Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V., 2006. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. Medical Imaging, IEEE Transactions on 25, 626 –639.

Studholme, C., Hill, D., Hawkes, D., 1999. An overlap invariant entropy measure of 3d medical image alignment . Pattern Recognition 32, 71–86.

Viola, P., Wells III, W., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vision 24, 137–154.

Wachinger, C., Navab, N., 2012. Entropy and laplacian images: Structural representations for multi-modal registration. Medical Image Analysis 16, 1 – 17.

Yi, Z., Soatto, S., 2011. Multimodal registration via spatial-context mutual information, in: Information Processing in Medical Imaging. Springer Berlin / Heidelberg. volume 6801, pp. 424–435.

Zhuang, X., Arridge, S., Hawkes, D., Ourselin, S., 2011. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. Medical Imaging, IEEE Transactions on 30, 1819 –1828.

Zikic, D., Baust, M., Kamen, A., Navab, N., 2010a. Generalization of deformable registration in riemannian sobolev spaces, in: Jiang, T., Navab, N., Pluim, J., Viergever, M. (Eds.), Medical Image Computing and Computer-Assisted Intervention MICCAI 2010. Springer Berlin / Heidelberg. volume 6362 of *Lecture Notes in Computer Science*, pp. 586–593.

Zikic, D., Kamen, A., Navab, N., 2010b. Revisiting horn and schunck: Interpretation as gauss-newton optimisation, in: BMVC, pp. 1–12.

Appendix A. Supplementary data

Appendix A.1. Parameters chosen for all compared methods

Table A.4: Parameter variation to obtain best results for Case 5 of CT dataset. Average target registration error (TRE) is given in mm (before registration TRE=7.10 mm). For values in brackets the transformation resulted in some negative Jacobians. Selected (fixed) settings are in bold.

Metric				
eSSD	Regularisation α	(0.1)	**0.5**	1.0
	TRE	(1.92)	2.06	2.22
	Gaussian patch σ	**1.0**	2.0	3.0
	TRE	2.06	2.19	2.23
	Histogram bins	16	**32**	(64)
	TRE	2.32	2.06	(1.98)
NMI	Regularisation α	(0.1)	**0.25**	0.5
	TRE	(2.36)	2.26	2.40
	Histogram bins	64	**128**	256
	TRE	2.26	2.26	2.25
CMI	Regularisation α	(0.01)	**0.05**	0.1
	TRE	(2.37)	2.37	2.69
	Histogram bins	64	128	**256**
	TRE	2.98	2.68	2.37
MIND	Regularisation α	(0.01)	**0.05**	0.1
	TRE	(2.02)	1.85	1.96
	Gaussian patch σ	**0.5**	1.0	1.5
	TRE	1.78	1.85	2.05
	Search region	**6-neighbour**	sparse	dense
	TRE	1.85	1.89	1.95

Table A.5: Parameters chosen for multi-modal 3D CT/MRI registration. To account for the increased noise and more complex intensity relations across modalities, we slightly increase the regularisation parameter α.

Metric		
eSSD	Regularisation α	0.5
	Gaussian patch σ	1.0
	Histogram bins	32
NMI	Regularisation α	0.25
	Histogram bins	128
CMI	Regularisation α	0.25
	Histogram bins	32
MIND	Regularisation α	0.1
	Gaussian patch σ	1.0
	Search region	6-neighbour