

Non-local shape descriptor: A new similarity metric for deformable multi-modal registration

Mattias P. Heinrich^{1,2,*}, Mark Jenkinson², Manav Bhushan^{1,2}, Tahreema Matin³, Fergus V. Gleeson³, Sir J. Michael Brady⁴, and Julia A. Schnabel¹

¹ Institute of Biomedical Engineering, University of Oxford, UK

² Oxford University Centre for Functional MRI of the Brain, UK

³ Department of Radiology Churchill Hospital, Oxford, UK

⁴ Department of Radiation Oncology and Biology, University of Oxford, UK

`mattias.heinrich@eng.ox.ac.uk`, `julia.schnabel@eng.ox.ac.uk`

`http://users.ox.ac.uk/~shil3388`

Abstract. Deformable registration of images obtained from different modalities remains a challenging task in medical image analysis. This paper addresses this problem and proposes a new similarity metric for multi-modal registration, the non-local shape descriptor. It aims to extract the shape of anatomical features in a non-local region. By utilizing the dense evaluation of shape descriptors, this new measure bridges the gap between intensity-based and geometric feature-based similarity criteria. Our new metric allows for accurate and reliable registration of clinical multi-modal datasets and is robust against the most considerable differences between modalities, such as non-functional intensity relations, different amounts of noise and non-uniform bias fields. The measure has been implemented in a non-rigid diffusion-regularized registration framework. It has been applied to synthetic test images and challenging clinical MRI and CT chest scans. Experimental results demonstrate its advantages over the most commonly used similarity metric - mutual information, and show improved alignment of anatomical landmarks.

1 Introduction

Advances in medical image registration techniques have resulted in a number of robust and accurate methods for deformable registration of scans of the same modality [1]. However, the registration of images from different modalities remains challenging. Alignment of multi-modal images helps to relate relevant information from different scans and to find the corresponding anatomical location of the response of functional imaging in structural scans. Intensity relations between those scans are not functional and can vary locally.

Mutual information (MI) is derived from information theory and measures the statistical dependency of two random variables. It was first introduced to

* We would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre. JAS also acknowledges funding from EPSRC EP/H050892/1.

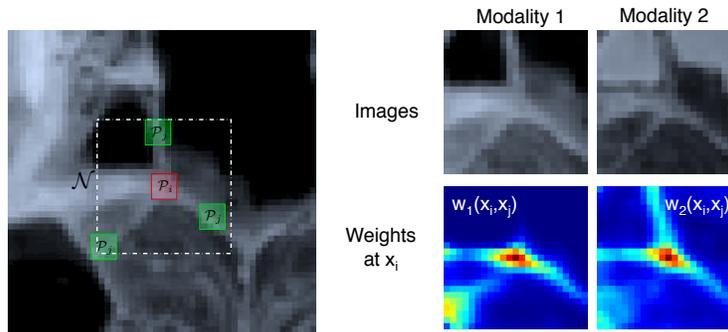


Fig. 1: Estimation of the non-local shape descriptor for the same feature at location \mathbf{x}_i in two different modalities, red and blue colour channels of cryosection (see text for further details).

medical image registration for the rigid alignment of multi-modal scans [2][3], and later used successfully in a variety of applications, including deformable registration. It is based on the assumption that a lower entropy of the joint intensity distribution corresponds to a better alignment. However, in several practical applications, additional constraints must be made or extensions added. Several weaknesses of MI for non-rigid registration have been identified [4]. For example, it is affected by non-uniform intensity distributions like bias fields. MI is intrinsically a global measure and therefore local deformations can lead to local minima in the solution as shown in [5]. To overcome these difficulties, we introduce a novel similarity metric for multi-modal image registration.

2 Non-local shape descriptor

We propose the non-local shape descriptor (NLSD), which defines a response related to the shape of image features at each location in both images to be registered. The shape descriptor is well adapted to medical image registration purposes, because it aims to extract anatomically meaningful geometric shapes.

The proposed similarity term is derived from a very efficient denoising technique, non-local means [6]. For the purpose of denoising it is necessary to find structural similarity in an extended non-local region of an image feature. The values of the most similar patches in the non-local search window contribute to a weighted average of the denoised central voxel. In this paper we will use the non-local weights to extract a geometric descriptor, which forms the basis of the proposed multi-modal similarity metric. A related descriptor, the self-similarity descriptor, has been presented for the application of object detection in [7].

We search for similar patches in a limited non-local region \mathcal{N} around the current voxel of interest \mathbf{x}_i . Within \mathcal{N} all patches \mathcal{P}_j are compared to the patch centred on \mathbf{x}_i . This concept is illustrated in Fig. 1, showing a magnification of an image feature in two different modalities, in this case the blue and red

colour channel of a cryosection. The dashed white line delimits the non-local search region, while the green squares outline exemplary patches \mathcal{P}_j , and the red square the central patch \mathcal{P}_i .

A weight $w(\mathbf{x}_i, \mathbf{x}_j)$ is assigned to each location \mathbf{x}_j in \mathcal{N} according to an exponentially decaying distance function based on the Euclidean distance between two patches \mathcal{P}_i and \mathcal{P}_j .

$$w(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\sum_{\Delta\mathbf{x}} \|I(\mathbf{x}_i + \Delta\mathbf{x}) - I(\mathbf{x}_j + \Delta\mathbf{x})\|^2}{\sqrt{2}\sigma^2}\right) \quad (1)$$

where $\Delta\mathbf{x}$ is defined over the range of voxels within a patch centred at $\mathbf{0}$. We thereby try to find patches within the non-local search region that are similar to the central patch. The value for σ^2 is the local variance of the noise and can be directly estimated from the 3D image data (see [8] for details).

The similarity metric at a given position \mathbf{x} is defined as the normalised cross correlation (NCC) of the respective weights w_1 and w_2 for the images I_1 and I_2 :

$$\text{NLSD}(\mathbf{x}) = \frac{\sum_k ((w_1(\mathbf{x}, \mathbf{x}_k) - \bar{w}_1) \cdot (w_2(\mathbf{x}, \mathbf{x}_k) - \bar{w}_2))}{\sqrt{\sum_k (w_1(\mathbf{x}, \mathbf{x}_k) - \bar{w}_1)^2} \sqrt{\sum_k (w_2(\mathbf{x}, \mathbf{x}_k) - \bar{w}_2)^2}}, k \in \mathcal{N}_{\mathbf{x}} \quad (2)$$

where \bar{w} is the mean of all weights within the non-local region $\mathcal{N}_{\mathbf{x}}$. NCC is robust against noise, but to accommodate for missing correspondences a mutual-saliency weighting [9] could potentially be beneficial.

We have implemented the calculation of this new similarity term as a convolution filter to evaluate the SSD of two patches. The pointwise product of both images is obtained and subsequently convolved with a uniform averaging filter. For the calculation of the weights within the non-local region \mathcal{N} the second image is shifted by $\mathbf{x}_k - \mathbf{x}_0 (\forall k \in \mathcal{N})$ and the averaging filter is applied again. This implementation speeds up the calculation of the similarity metric substantially and avoids the need for preselection of potentially good weights as proposed in [8]. The size of the non-local region should be as large as possible, but for practical applications a size of 15x15 for 2D experiments and 7x7x7 for 3D images together with a patch size of 3x3 and 3x3x3, respectively, has been found to be sufficient to obtain good responses for the shape descriptor.

3 Registration framework

Within the non-rigid registration, we aim to minimize the following cost function w.r.t. 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 regularization term:

$$\underset{\mathbf{u}}{\text{argmin}} = \int_{\Omega} \mathcal{S}(I_1(\mathbf{x}), I_2(\mathbf{x} + \mathbf{u}))^2 + \alpha \text{tr}(\nabla\mathbf{u}(\mathbf{x})^T \nabla\mathbf{u}(\mathbf{x}))^2 d\mathbf{x} \quad (3)$$

Simple gradient descent methods show slow convergence, especially in homogeneous regions [10]. Since the objective function to be minimized is of the form of $\sum_i f_i^2$, we can apply the Gauss-Newton optimization method, where f is

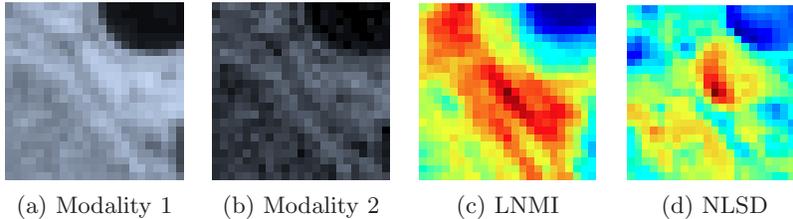


Fig. 2: Feature location in both images ((a) red and (b) blue channel of colour cryosection). Response in search window of (c) LNMI and (d) NLSD. Our proposed method shows a more discriminative peak in the centre.

minimized iteratively with the update rule: $(\mathbf{J}^T \mathbf{J}) \mathbf{u}_{\text{gn}} = -\mathbf{J}^T f$, where \mathbf{J} is the derivative of f w.r.t. \mathbf{u} . This can be adapted to our regularized 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 regularization term is linear w.r.t. \mathbf{u} as the differential operator is linear. The resulting update step given an initial or previous deformation field \mathbf{u}_{prev} is given by:

$$(\nabla \mathcal{S}^T \nabla \mathcal{S} + \alpha \Delta) \mathbf{u}_{\text{gn}} = -(\nabla \mathcal{S}^T \mathcal{S} + \alpha \Delta \mathbf{u}_{\text{prev}}) \quad (4)$$

Equation 4 is solved using 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 regularizer. We set $\alpha = 1$ in all experiments.

We have also implemented MI, the classic choice of a multi-modal similarity criterion, within the same deformable registration framework. For the variational optimisation we need to evaluate the similarity function at each location, therefore a local derivation of mutual information is used, as described in [11]. An overview of the possibilities of variational implementations other statistical similarity terms, as well as a discussion of a locally weighted computation of the global measures, is given in [12].

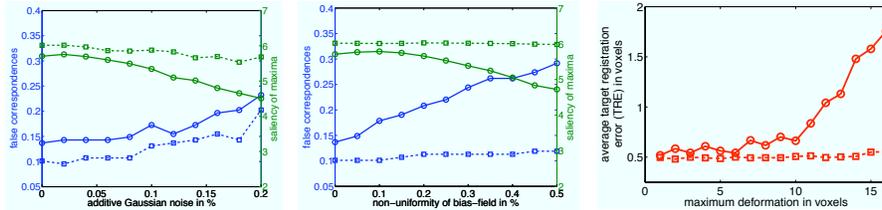
Given the joint probability $p_{12}(\mathbf{i})$ of the co-occurrence of an intensity pair $\mathbf{i} = (i_1, i_2)^T$ in two images I_1 and I_2 and the two marginal intensity probabilities $p_1(i_1)$ and $p_2(i_2)$, local normalised mutual information (LNMI) at location \mathbf{x} is defined as (using the global entropy of I_1 for normalization):

$$\text{LNMI}(\mathbf{x}) = \log \left(\frac{p_{12}(I_1(\mathbf{x}), I_2(\mathbf{x}))}{p_1(I_1(\mathbf{x})) \cdot p_2(I_2(\mathbf{x}))} \right) \frac{1}{\int_{\mathbf{x}} p_1(I_1(\mathbf{x})) \log(p_1(I_1(\mathbf{x}))) d\mathbf{x}} \quad (5)$$

The joint and marginal histograms are recalculated at each iteration and smoothed with a Parzen window kernel of size 5×5 with a standard deviation of 0.5. We use 128 histogram bins.

4 Saliency and robustness of correspondences

We examine the ability of our new similarity metric to distinguish between anatomical features in different modalities under the influence of local defor-



(a) Influence of increasing noise. (b) Influence of increasing bias field. (c) TRE with increasing strength of deformations.

Fig. 3: (a,b) Saliency and robustness of both similarity metrics are compared. (c) TRE of registrations of synthetic deformations (see text for details). LNMI is displayed with solid lines and circles, NLSD with dashed lines and squares.

mations, additive noise and non-uniform bias fields. A ground truth is provided by using two images of different colour channels taken from a cryosection of the *Visible Human Project*, which are intrinsically aligned. A number of landmarks were selected at the same location in both images using the Harris corner detector. The similarity metric was then calculated between a point in the first image and all locations in the second image within a window of 23×23 around that location. Figure 2 shows one selected point and the search window in both images. The proposed metric can better distinguish the local maximum in the centre.

We run these comparisons over all feature locations and quantify the results using two criteria. First, the distance of the maxima of the similarity function is compared with the ground truth location. Deviations of more than 3 pixels are counted as false correspondences. We define the fractional amount of false matches as robustness. Second, the saliency or discrimination between the maximum and its surrounding values is quantified by convolving the similarity response with a Mexican hat function ($\sigma = 1$), so that high positive values are characteristic for a high saliency in the similarity function (see Fig. 3 (a,b)).

5 Results

We performed registrations, using a multi-resolution scheme, for the synthetic test images on which a simulated deformation was applied. The deformations of varying strengths are obtained using a uniform B-spline grid and random control point displacements. In Fig. 3 (c) the average target registration error (TRE) between ground truth deformations and registration are compared for LNMI and NLSD for increasing magnitudes of deformations. For larger deformations, NLSD shows higher accuracy, and the TRE remains almost unaffected, while for LNMI the TRE strongly deteriorates. This demonstrates the disadvantages of mutual information, when the initial estimate of the joint intensity distribution is not close enough to the real distribution, due to large deformations, and the similarity term is susceptible to local optima.

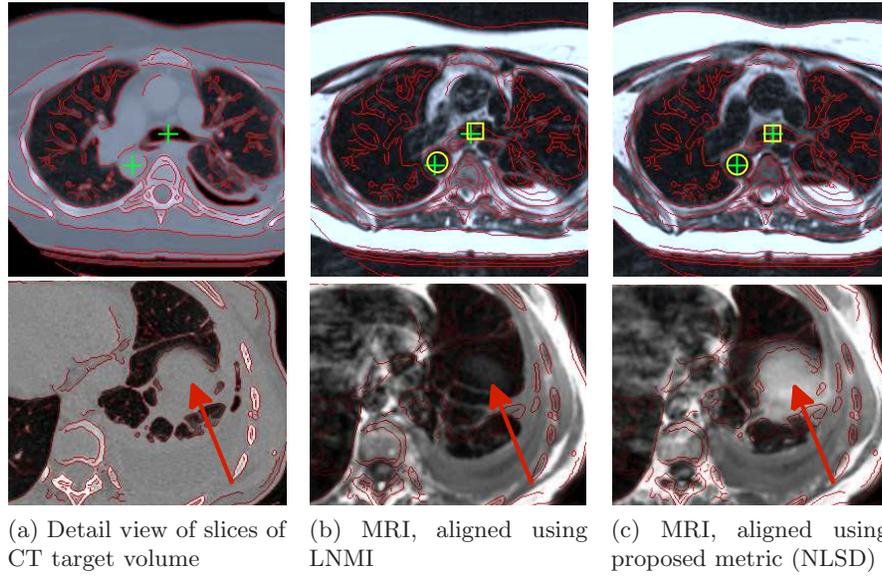


Fig. 4: Axial slices through CT and MR image of the lungs of two patients with empyema. Contours of CT are shown for visual guidance. Landmarks in top row: descending aorta (\odot) and carina (\square) after non-rigid registration compared to the gold standard (+) demonstrate a better alignment for NLSD. The example below shows a substantial improvement for the dome of the diaphragm (arrow).

We then applied our proposed technique to a clinical dataset of 11 patients, which were scanned with both CT and MRI. All patients suffered from empyema, a lung disease where the pleura gets infected and excess fluid fills up the pleural space. This causes the lung to collapse and the extra fluid turns into an abscess. 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 make it difficult to relate the scans for the clinician. A particular challenge for the registration are large slice thicknesses of up to 8 mm used for the MRI acquisition. The background is removed using a threshold and a morphological filter. For the registration, first a rigid body transformation is estimated using a blockmatching algorithm [13]. In the second step, the proposed non-rigid registration is performed, using a multiresolution scheme with 3 levels. Similarity terms and their derivatives are recalculated before each iteration of the Gauss-Newton optimisation method. The iterations are stopped when the mean of the cost function does not further decrease.

We compare the results of our proposed metric, non-local shape descriptors (NLSD), against local normalised mutual information (LNMI). The running time for one 3D registration of images with a size of $253 \times 253 \times 132$ voxels is 43 minutes for NLSD and 24 minutes for LNMI on a single core. The range of values for the determinant of the Jacobian of the deformation fields are $[0.25, 1.81]$ for NLSD

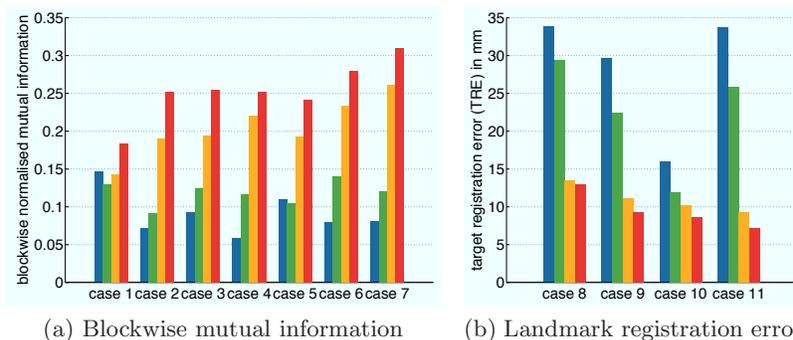


Fig. 5: Quantitative evaluation of registration outcome: ■ before registration, ■ rigid-body alignment, ■ non-rigid registration using LNMI, and ■ using the proposed metric NLSD.

and $[0.26, 1.81]$ for LNMI, thus no physically implausible folding occurred and all transformations are invertible.

The registration outcome for two cases is displayed in Fig. 4. The target CT volume is shown along with the aligned MR images with deformable registration using LNMI and our proposed metric. Both the contours of the CT and manual anatomical landmarks reveal the advantages and improved accuracy of the presented method. To compare our findings quantitatively, we first calculated the intensity-based similarity before and after registration using the presented metrics. Although an improvement of a similarity function does not necessarily ensure anatomical correspondence, it can highlight differences between methods. We use mutual information calculated within cubic blocks of 30^3 voxels to reduce the influence of the non-uniform bias field in the MRI scans. Figure 5 (a) shows an improvement of this measure using NLSD, for all seven cases, over MI. Additionally a clinical expert manually selected landmarks for the four remaining cases. Between 12 and 15 corresponding landmarks were selected in the four 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 (motion artifacts) and changes of the pathology in the diseased areas between scans. On average, the target registration error (TRE) could be further reduced by about 2 mm using our new metric compared to LNMI (see Fig. 5 (b)).

6 Conclusion

In this work a new similarity metric for deformable multi-modal registration is proposed. The non-local shape descriptor (NLSD) aims to extract the most descriptive geometric features in medical images, while being nearly independent of non-functional intensity relations, non-uniform intensity fields and additive noise. This new metric can robustly find correspondences in different modalities

and strongly discriminate between salient points. The technique is implemented in a variational, diffusion-regularized registration framework and compared to the most commonly used alternative - mutual information. We demonstrate that NLSD achieves much improved and more accurate registration results, especially in the case of large deformations. We validate our findings for the application of deformable registration of clinical MR and CT scans of diseased patients. Anatomical landmarks chosen by an expert clinician show improved alignment using our metric. A more thorough evaluation, including more landmarks and a comparison within different transformation models, will be addressed in the future.

References

1. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Pluim, J.P.W.: Evaluation of Methods for Pulmonary Image Registration: The EMPIRE10 Study. In: *Medical Image Analysis for the Clinic: A Grand Challenge*. (2010)
2. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information. *IEEE Trans. Med. Imaging* 16(2), 187–198 (1997)
3. Viola, P., Wells III, W.M.: Alignment by Maximization of Mutual Information. *Int. J. Comput. Vision* 24(2), 137–154 (1997)
4. Haber, E., Modersitzki, J.: Intensity Gradient Based Registration and Fusion of Multi-modal Images. *Methods Inf. Med.* 46(3), 292–299 (2007)
5. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid Image Registration Using Conditional Mutual Information. *IEEE Trans. Med. Imaging* 29(1), 19–29 (2010)
6. Buades, A., Coll, B., Morel, J.M.: A Non-Local Algorithm for Image Denoising. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 60–65. IEEE Computer Society (2005)
7. Shechtman, E., Irani, M.: Matching Local Self-Similarities across Images and Videos. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE Computer Society, Los Alamitos (2007)
8. Coupé, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *IEEE Trans. Med. Imaging* 27(4), 425–441 (2008)
9. Ou, Y., Davatzikos, C.: DRAMMS: Deformable Registration via Attribute Matching and Mutual-Saliency Weighting. In: Prince, J., Pham, D., Myers, K. (eds.) *IPMI 2009*. LNCS, vol. 5636, pp. 50–62. Springer, Heidelberg (2009)
10. Zikic, D., Baust, M., Kamen, A., Navab, N.: Generalization of Deformable Registration in Riemannian Sobolev Spaces. In: Jiang, T., Navab N., Pluim, J., Viergever, M. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 586–593. Springer, Heidelberg (2010)
11. Rogelj, P., Kovačič, S., Gee, J.C.: Point similarity measures for non-rigid registration of multi-modal data. *Comput. Vis. Image Und.* 92(1), 112–140 (2003)
12. Hermosillo, G., Chéfd’hotel, C., Faugeras, O.: Variational Methods for Multimodal Image Matching. *Int. J. Comput. Vision* 50(3), 329–343 (2002)
13. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In: Delp, S., DiGoia, A., Jaramaz, B. (eds.) *MICCAI 2000*. LNCS, vol. 1935, pp. 373–380. Springer, Heidelberg (2000)