

Towards Realtime Multimodal Fusion for Image-Guided Interventions using Self-Similarities

Mattias P. Heinrich^{1,2}, Mark Jenkinson², Bartłomiej W. Papież¹,
Sir Michael Brady³, and Julia A. Schnabel¹

¹ Institute of Biomedical Engineering,

Department of Engineering, University of Oxford, UK

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

³ Department of Oncology, University of Oxford, UK

mattias.heinrich@eng.ox.ac.uk, <http://users.ox.ac.uk/~shil3388>

Abstract. Image-guided interventions often rely on deformable multi-modal registration to align pre-treatment and intra-operative scans. There are a number of requirements for automated image registration for this task, such as a robust similarity metric for scans of different modalities with different noise distributions and contrast, an efficient optimisation of the cost function to enable fast registration for this time-sensitive application, and an insensitive choice of registration parameters to avoid delays in practical clinical use. In this work, we build upon the concept of structural image representation for multi-modal similarity. Discriminative descriptors are densely extracted for the multi-modal scans based on the “self-similarity context”. An efficient quantised representation is derived that enables very fast computation of point-wise distances between descriptors. A symmetric multi-scale discrete optimisation with diffusion regularisation is used to find smooth transformations. The method is evaluated for the registration of 3D ultrasound and MRI brain scans for neurosurgery and demonstrates a significantly reduced registration error (on average 2.1 mm) compared to commonly used similarity metrics and computation times of less than 30 seconds per 3D registration.

Keywords: multimodal similarity, discrete optimisation, neurosurgery

1 Introduction

Deformable multi-modal registration plays an important role for image-guided interventions, where scans are often acquired using different modalities, e.g. to propagate segmentation information for image-guided radiotherapy [9]. The alignment of multi-modal scans is difficult, because there can be a large amount of motion between scans, the intra-operative scan is often of lower scan quality than diagnostic scans and no functional relationship between intensities across modalities exists. In this work, we address the registration of intra-operative 3D ultrasound (US) to pre-operative magnetic resonance imaging (MRI) for image-guided neurosurgery. The brain tissue exhibits non-rigid deformations after opening the skull (*brain shift* [11]), which needs to be compensated to relate

the intra-operative ultrasound to the MRI scan (which has a higher quality and can give a better guidance for tumour resection).

2 Previous Work

Mutual information (MI) has been frequently used in rigid multi-modal registration [10, 17]. However, for deformable multi-modal registration, many disadvantages have been identified for MI-based similarity measures [13]. In particular, the locally varying noise distribution and speckle pattern of ultrasound make the estimation of a global intensity mapping difficult. The results in [14] suggest that using standard MI is insufficient for US-MRI registration. The authors propose the use of a multi-feature α -MI metric as presented in [16]. This, however, comes at a great computational cost, which is a disadvantage for this time sensitive application. Another possible approach is to use a simulated ultrasound-like image based on an MRI [1] or computed tomography (CT) scan [19] for registration. An accurate ultrasound simulation is, however, far from trivial, especially for interventional scans.

For these reasons, structural image representations have gained great interest for deformable multi-modal registration. The motivation is that, once the images are transformed into a representation independent of the underlying image acquisition, efficient monomodal optimisation techniques can be employed. In [18], a scalar structural representation based on local entropy has been successfully applied to the deformable registration of different MRI modalities and the rigid alignment of MRI and CT brain scans. De Nigris et al. [2] used gradient orientation to drive a rigid multi-modal registration of brain scans. In [4], we proposed a multi-dimensional “modality independent neighbourhood descriptor” (MIND) based on the concept of local self-similarities (LSS) [15] and applied it to the non-rigid registration of CT and MRI chest scans. Self-similarities were also employed in [14], however, not as a structural representation, but instead using the Earth Mover’s Distance as self-similarity distance within the multi-dimensional feature-space of α -MI. There are a number of limitations of these approaches. First, scalar representations [18] are often not sufficiently discriminative to drive a non-rigid registration with many degrees of freedom. Second, high-dimensional descriptors are often computationally too expensive for the use in interventional applications. Third, the inherent strong noise and imaging artefacts of US make a robust estimation of structural image representations challenging.

In this work, we address these challenges by introducing a novel contextual image descriptor: the “self-similarity context” (SSC). The motivation and derivation of SSC is presented in Sec. 3.1. A descriptor quantisation and an efficient point-wise distance metric are proposed. In Sec. 3.2 the employed discrete optimisation framework is described and a simple approach is presented, which ensures symmetric transformations. Quantitative experiments of a neurosurgical application are presented in Sec. 4 for an evaluation of our approach and a comparison to other similarity metrics and previously published methods on the same dataset, based on manual landmark correspondences.

3 Methods

SSC is estimated based on patch-based self-similarities in a similar way as e.g. LSS or MIND, but rather than extracting a representation of local shape or geometry, it aims to find the context around the voxel of interest. Therefore, the negative influence of noisy patches can be greatly reduced, making this approach very suitable for ultrasound registration. Spatial context has been successfully applied to object detection [7] and is the driving force of pictorial structures [3].

3.1 Self-similarity context

Self-similarity can be described by a distance function between image patches within one image I (sum of squared differences SSD can be used within the same scan), a local or global noise estimate σ^2 , and a certain neighbourhood layout \mathcal{N} for which self-similarities are calculated. For a patch centred at \mathbf{x} , the self-similarity descriptor $\mathcal{S}(I, \mathbf{x}, \mathbf{y})$ is given by:

$$\mathcal{S}(I, \mathbf{x}, \mathbf{y}) = \exp\left(-\frac{SSD(\mathbf{x}, \mathbf{y})}{\sigma^2}\right) \quad \mathbf{x}, \mathbf{y} \in \mathcal{N} \quad (1)$$

where \mathbf{y} defines the centre location of a patch within \mathcal{N} . In [4] and [15], the neighbourhood layout was defined to always include the patch centred around \mathbf{x} for pairwise distance calculations. This has the disadvantage that image artefacts or noise within the central patch always have a direct adverse effect on the self-similarity descriptor. We therefore propose to completely avoid using the patch at \mathbf{x} for the calculation of the descriptor for this location. Instead, all pairwise distances of patches within the six neighbourhood (with a Euclidean distance of $\sqrt{2}$ between them) are used. The spatial layout of this approach is visualised in Fig. 1 and compared to MIND (showing the central patch in red, patches in \mathcal{N} in grey, and edges connecting pairs for which distances are calculated in blue). The aim of SSC is not to find a good structural representation of the underlying shape, but rather the context within its neighbourhood. The noise estimate σ^2 in Eq. 1 is defined to be the mean of all patch distances. Descriptors are normalised so that $\max(\mathcal{S}) = 1$.

Pointwise Multimodal Similarity using SSC: Structural image representations are appealing, because they enable multi-modal registration using simple similarity metric across modalities. Once the descriptors are extracted for both images, yielding a vector for each voxel, the similarity metric between locations \mathbf{x}_i and \mathbf{x}_j in two images I and J can be defined as the sum of absolute differences (SAD) between their corresponding descriptors. The distance D between two descriptors is therefore:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{|\mathcal{N}|} \sum_{\mathbf{y} \in \mathcal{N}} |\mathcal{S}(I, \mathbf{x}_i, \mathbf{y}) - \mathcal{S}(J, \mathbf{x}_j, \mathbf{y})| \quad (2)$$

Equation 2 requires $|\mathcal{N}|$ computations to evaluate the similarity at one voxel. Discrete optimisation frameworks, as the one employed here, use many cost function evaluations per voxel. In order to speed up the computations, we propose to

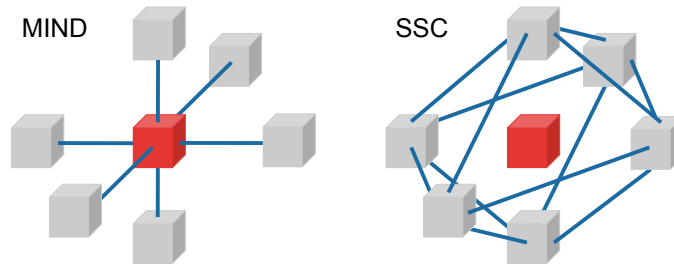


Fig. 1. Concept of self-similarity context (SSC) compared to MIND with six-neighbourhood (6-NH). The patch around the voxel of interest is shown in red, all patches within its immediate 6-NH are shown in grey. Left: All patch distances (shown with blue lines) used for MIND within the 6-NH take the centre patch into account. Right: Geometrical and structural context can be better described by SSC using all patch-to-patch distances, none of which is dependent on the central patch.

quantise the descriptor to a single integer value with 64 bits, without significant loss of accuracy. The exact similarity evaluation of Eq. 2 can then be obtained using the Hamming distance between two descriptors using only one operation per voxel. A descriptor using self-similarity context consists of 12 elements, for which we use 5 bits per element, which translates into 6 different possible values (note that we cannot use a quantisation of 2^5 because the Hamming weight only counts the number of bits, which differ). Figure 2 illustrates the concept, which could also be employed for other multi-feature based registration techniques.

3.2 Discrete Optimisation

Discrete optimisation is used in this work, because it is computationally efficient, no derivative of the similarity cost is needed, local minima are avoided and large deformations can be covered by defining an appropriate range of displacements \mathbf{u} . Here, we adopt our recent approach [5, 6], which uses a block-wise parametric transformation model with belief propagation on a tree graph (BP-T) [3]. The regularisation term penalises squared differences of displacements for neighbouring control points and is weighted with a constant λ . Not all similarity terms within the influence region of each control point, but only a randomly chosen subset of them are taken into account. This concept, which has also been used in stochastic gradient optimisation [8] and in [2], greatly reduces the computation of the similarity term without loss of accuracy (as shown in [6]).

Inverse Consistent Transformations: We introduce a simple scheme to obtain inverse consistent mappings, given the forward and backward displacement fields \mathbf{u}^n and \mathbf{v}^n respectively (which are independently calculated), by iteratively updating the following equations:

$$\begin{aligned} \mathbf{u}^{n+1} &= 0.5(\mathbf{u}^n - \mathbf{v}^n(\mathbf{x} + \mathbf{u}^n)) \\ \mathbf{v}^{n+1} &= 0.5(\mathbf{v}^n - \mathbf{u}^n(\mathbf{x} + \mathbf{v}^n)) \end{aligned} \quad (3)$$

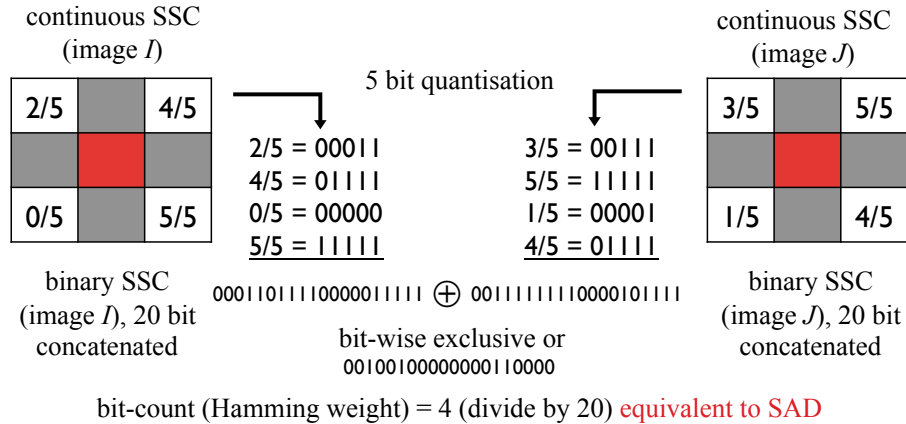


Fig. 2. Concept of using the Hamming distance to speed up similarity evaluations. Continuous valued descriptor entries (here: $|\mathcal{N}| = 4$) are quantised to a fixed number of bits and concatenated. The similarity of Eq. 2 can then be evaluated using only one bitwise XOR and bit-count operation.

4 Registration of US and MRI Scans for Neurosurgery

We applied the presented approach to a set of 13 pairs of pre-operative MRI and pre-resection 3D ultrasound (US) images of the Brain Images of Tumours for Evaluation (BITE) database [12] from the Montreal Neurological Institute[†]. The MRI scans have an isotropic resolution of 0.5 mm and the US are resampled to the same resolution (the MRI scans are then cropped to have the same dimensions and a similar field of view). Roughly 27 corresponding anatomical landmarks have been selected for each scan pair by a neurosurgeon and two experts. The same dataset was recently used by [14], and therefore enables a direct comparison. They apply multi-feature α -MI [16] with a stochastic gradient descent optimisation [8], and extend this framework using a self-similarity weighting within the feature space, calling the new metric SeSaMI.

We use the following parameter settings for the discrete optimisation: three scales of control point spacings of $\{6, 5, 4\}$ mm, 50 similarity term samples per control point and a dense displacement search range of $\{12, 5, 2\}$ mm (with a spacing of $\{2, 1, 0.5\}$ mm). This corresponds to roughly 10^7 degrees of freedom for the optimisation. We employ three different similarity metrics: blockwise MI, MIND [4] and SSC. For the self-similarity calculations a patch-size of $3 \times 3 \times 3$ voxels and a distance between neighbouring patches of 2 voxels was chosen. Blockwise MI is computed the same way as global MI [10], but a new joint histogram is estimated for each control point and each displacement using 100 samples within the cubic influence region of the control point. We empirically found that 8 histogram bins give best results together with a Parzen window

[†] Publicly available at www.bic.mni.mcgill.ca/Services/ServicesBITE

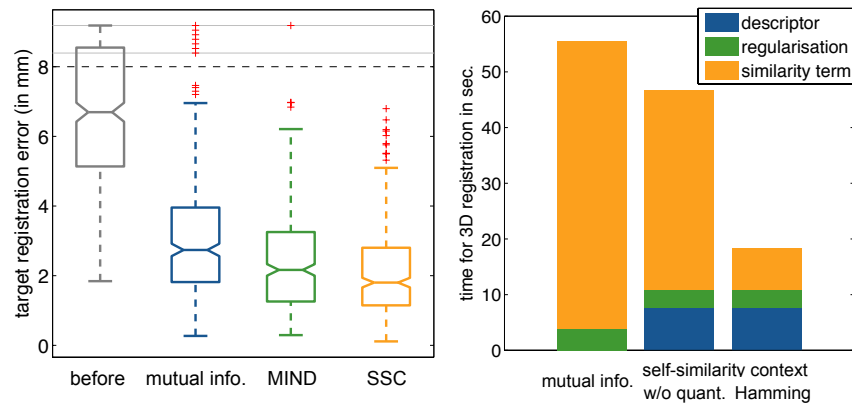


Fig. 3. Deformable registration of 13 cases of MRI-US brain scans, evaluated with ≈ 27 expert landmarks per case. The registration error of SSC (2.12 ± 1.29 mm) is significantly lower than that of MIND (2.45 ± 1.44 mm) and mutual information (3.07 ± 1.74 mm) within the same discrete optimisation framework ($p = 7 \times 10^{-4}$ and $p < 10^{-6}$ respectively). The computation time per registration using SSC and the Hamming distance (≈ 20 sec per 3D pair) is more than twice as fast as MI and SSC without quantisation. The use of α -MI with a continuous optimisation approach takes about 120 min. [14], which is too long for this time-concerning application.

smoothing with $\sigma_P = 0.5$. An optimal regularisation parameter $\lambda = 0.5$ was found for MIND and SSC, and $\lambda = 0.25$ for MI. All resulting transformations are free from singularities (invertible) with an average complexity measured as standard deviation of the Jacobian of 0.08.

Quantitative Results: The average initial target registration error (TRE) in our experiments is 6.76 ± 2.20 mm (this value is only 4.12 mm in [14], because additional US-tracking information is used). SSC achieves the best overall registration accuracy of 2.12 ± 1.29 mm (see Fig. 3), which is a significant improvement compared to MIND (2.45 ± 1.44 mm) and mutual information (3.07 ± 1.74 mm). The much more complex α -MI metric and its SeSaMI variant used in [14] yield a higher TRE of 2.50 and 2.34 mm, respectively. The computation time of SSC (20 sec) is smaller than using MI (55 sec) within the same discrete optimisation framework and much smaller than using α -MI (120 min). Additionally, when using SSC, the setting of the regularisation parameter λ has a low sensitivity (less compared to MIND and MI) causing an increase of TRE of only 0.1 mm when choosing $4 \times$ larger or smaller values. Figure 4 shows an example of the registration problem and the resulting alignment using SSC.

5 Conclusion

This paper addresses the challenging deformable registration of pre-operative MRI to intra-operative ultrasound for neurosurgery. A novel image descriptor

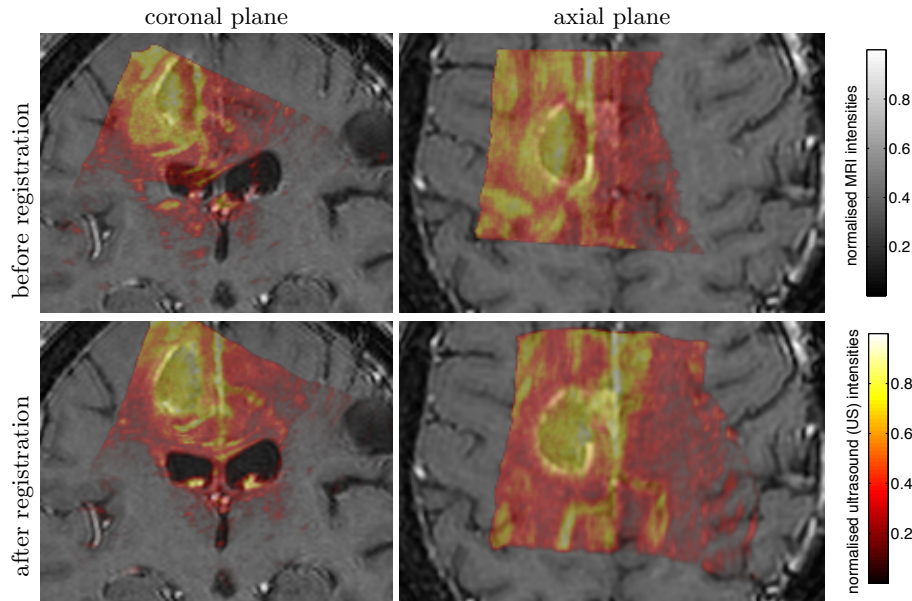


Fig. 4. Deformable MRI-US registration results of BITE dataset using SSC with discrete optimisation. The intra-operative ultrasound scan is shown as false colour overlay over the grayscale MRI intensities. A clearly improved alignment of the ventricles and the solid tumour is visible after registration (bottom row).

the “self-similarity context” (SSC) is presented, with low sensitivity to image noise, and a quantisation scheme for fast distance evaluations using Hamming weights. When used in a discrete optimisation framework with a stochastic similarity term sampling, a computation time of less than half a minute is achieved on a standard CPU and state-of-the-art registration accuracy with an average error of 2.12 mm, which is a statistically significant improvement over previous self-similarity based metrics [4] and mutual information. In the future, we plan a GPU implementation (which could then lead to real-time performance) and further comparisons to other structural image representations (e.g. gradient orientation [2]) and the application of our approach to further applications of image-guided interventions.

Acknowledgements: We would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre.

References

1. Arbel, T., Morandi, X., Comeau, R., Louis Collins, D.: Automatic non-linear MRI-ultrasound registration for the correction of intra-operative brain deformations. In: Niessen, W., Viergever, M. (eds.) MICCAI 2001. pp. 913–922. LNCS, Springer (2001)

2. De Nigris, D., Collins, D., Arbel, T.: Multi-modal image registration based on gradient orientations of minimal uncertainty. *IEEE Trans. Med. Imag.* (2012)
3. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. *Int. J. Comp. Vis.* 61, 55–79 (2005)
4. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, S.M., Schnabel, J.A.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Med. Imag. Anal.* 16(7), 1423 – 1435 (2012)
5. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: Delingette, H., Golland, P., Kensaku, M., Ayache, N. (eds.) *MICCAI 2012*, pp. 115–122. LNCS, Springer (2012)
6. Heinrich, M., Jenkinson, M., Brady, M., Schnabel, J.: MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans. Med. Imag.* (2013)
7. Heitz, G., Koller, D.: Learning spatial context: Using stuff to find things. In: Forsyth, D.A., Torr, P.H.S., Zisserman, A. (eds.) *ECCV 2008*, LNCS, vol. 5302, pp. 30–43. Springer (2008)
8. Klein, S., Staring, M., Pluim, J.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-Splines. *IEEE Trans. Image Proc.* 16(12), 2879 –2890 (2007)
9. Lu, C., Chelikani, S., Papademetris, X., Knisely, J.P., Milosevic, M.F., Chen, Z., Jaffray, D.A., Staib, L.H., Duncan, J.S.: An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotherapy. *Med. Imag. Anal.* 15(5), 772–785 (2011)
10. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* 16(2), 187–198 (1997)
11. Maurer, C.R., Hill, D.L.G., Martin, A.J., Liu, H., McCue, M., Rueckert, D., Lloret, D., Hall, W.A., Maxwell, R.E., Hawkes, D.J., Truwit, C.L.: Investigation of intraoperative brain deformation using a 1.5 Tesla interventional MR system: Preliminary results. *IEEE Trans. Med. Imaging* 17(5), 817–825 (1998)
12. Mercier, L., Del Maestro, R., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.: Online database of clinical MR and ultrasound images of brain tumors. *Med. Phys.* 39, 3253 (2012)
13. Pluim, J., Maintz, J., Viergever, M.: Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imag.* 19(8), 809–814 (2000)
14. Rivaz, H., Collins, D.: Self-similarity weighted mutual information: A new nonrigid image registration metric. In: Delingette, H., Golland, P., Kensaku, M., Ayache, N. (eds.) *MICCAI 2012*, pp. 91–98. LNCS, Springer (2012)
15. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: *CVPR 2007*. pp. 1 –8. IEEE (2007)
16. Staring, M., van der Heide, U., Klein, S., Viergever, M., Pluim, J.: Registration of cervical MRI using multifeature mutual information. *IEEE Trans. Med. Imag.* 28(9), 1412–1421 (2009)
17. Viola, P., Wells III, W.: Alignment by maximization of mutual information. *Int. J. Comput. Vision* 24(2), 137–154 (1997)
18. Wachinger, C., Navab, N.: Entropy and laplacian images: Structural representations for multi-modal registration. *Med. Imag. Anal.* 16(1), 1 – 17 (2012)
19. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Med. Imag. Anal.* 12(5), 577 (2008)