Robust Super-Resolution Reconstruction with Multi-Modal Registration and Guidance

Mattias P. Heinrich[1,2]
mattias.heinrich@eng.ox.ac.uk
Mark Jenkinson[2]
Sir Michael Brady[3]
Julia A. Schnabel[1]

[1] Institute of Biomedical Engineering
University of Oxford, UK
[2] Oxford University Centre for Functional MRI of the Brain, Oxford, UK
[3] Department of Oncology
University of Oxford, UK

Abstract

Increasingly, clinical practice involves acquiring multiple scans with different modalities for diagnostic tasks, as well as longitudinal studies to monitor patient response to therapy and for disease progression. Often, one of these scans is of substantially lower spatial resolution. In order to maintain the highest possible resolution of all scans for subsequent analysis steps, upsampling of the images is needed. Recently, new approaches have been proposed to make use of an inter-modality high-resolution image as a guidance for this upsampling process. While these techniques achieve a significantly higher quality for two perfectly aligned volumes compared to traditional interpolation methods, they are not robust against misalignments between the high- and low-resolution scans. We address this problem by incorporating a deformable multi-modal registration step in the super-resolution reconstruction process. We demonstrate an improved performance of our method compared to two different intra-modality interpolation-based techniques and an inter-modality guided approach without incorporation of registration.

1 Introduction and Background

Medical imaging modalities provide complementary information, due to their different physical acquisition principles. For example, computed tomography (CT) scans have high spatial resolution and good dense tissue contrast, while magnetic resonance images (MRI) excel in higher soft tissue contrast at the cost of a longer acquisition time and lower resolution. The acquisition of scans of multiple modalities is becoming clinical practice. Using their complementary information in the highest possible resolution is therefore of great clinical benefit, for example for pulmonary image analysis of thoracic images, where MRI acquisition is difficult to the breathing motion.

There are two principal challenges involved in the fusion of such multi-modal scans. First, patient motion and imaging distortion, which may have occurred between two scans, has to be compensated for. Rigid and deformable image registration has been widely studied to address this problem. Second, the inherently different spatial resolutions of the different modalities have to be reconciled to bring both scans into the same coordinate space. Conventionally, interpolation techniques are used to resample the lower resolution (LR) scan to match the higher-resolution (HR) scan (upsampling). However, such an approach uses only the neighbouring intensity values of the same low-resolution scan and ignores the available high-resolution information of the complementary scan; this in turn causes blurring of anatomical structures.

Recently, a number of methods have been proposed which use an inter-modality high-resolution image as a guide for the upsampling process. Kopf et al. [6] introduced a joint bilateral filtering approach for super-resolution (SR) reconstruction based on low-resolution exposure maps, chrominance image and stereo depth maps together with a high-resolution gray-scale image for guidance. In [7] and [8] super-resolution images of low-resolution T2-weighted MRI brain scans are obtained using a non-local weighting process that uses a high-resolution T1-weighted scan as prior. These methods all have in common that the perfect alignment of LR and HR images is required to obtain accurate and robust results. In clinical practice, this assumption generally does not hold true due to residual registration errors, particularly in the case of significantly different voxel dimensions.

We have developed a novel approach, which incorporates the estimation of a deformation field between the multimodal LR and HR scans into the super-resolution reconstruction process. Based on a point-wise multi-modal similarity metric and a diffusion-regularised Gauss-Newton optimisation, a deformation field is computed between the voxels in the LR and HR guidance scans. The deformation field is then used to transform the appropriate non-local weighting of the HR scan in the SR reconstruction. The process is iteratively refined, while the solution is constrained by the underlying imaging physics. Our approach is explained in detail in the next sections. We demonstrate its improved performance on clinically relevant tasks for SR reconstruction, including upsampling of chest MRI images using a HR CT scan as a guide.

2 Guided Upsampling with an Inter-Modality Prior

The observation model of the formation of a LR image $\mathbf{y}$ based on the HR image $\mathbf{x}$ is (as defined in [8]):

$$\mathbf{y} = DBW\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ represents the noise, D is a sub-sampling matrix, B is a blur operator, and W is the geometric transformation between the HR and LR images. In this work, we assume the noise $\mathbf{n}$ has been removed as a preprocessing step. Alternatively, an appropriate noise-model could be incorporated into the upsampling process. In the case of MRI scanners, the blur operator B can be approximated by a 3D boxcar function.

To ensure that the reconstructed SR volume $\hat{\mathbf{x}}$ complies with the physical acquisition model, an additional constraint, the subsampling consistency (SSC), was introduced by [2]. It requires the downsampled SR image to be equivalent to the original LR image:

$$\mathbf{y} - DBW\hat{\mathbf{x}} = 0 \tag{2}$$

The reconstruction step of our method follows the approach of [7]. Based on the assumption that similar voxels in the inter-modality HR image $\mathbf{z}$ are similar in the reconstructed SR image, we perform a weighted averaging using a joint filtering approach (see e.g. [6]). The weights of this filter are estimated from the non-local neighbourhood Ω of the voxels in the HR guidance image $\mathbf{z}$; but the averaging is performed on the reconstructed SR voxels. In contrast to linear invariant filters (e.g. Gaussian, Laplacian), the filter characteristic is data-dependent and therefore different for each voxel. Based on a previous estimate $\hat{\mathbf{x}}^t$ the intensities of the SR image can be calculated with the following equation:

$$\hat{\mathbf{x}}^{t+1}(x) = \frac{1}{c} \sum_{r \in \Omega} \hat{\mathbf{x}}^t(x+r) \exp^{-||\mathbf{z}(x)-\mathbf{z}(x+r)||^2/h} \tag{3}$$

where the $\Omega = \{0, \pm 1, \ldots, \pm r_{\max}\}^3$ defines 3D the non-local search region, c is a normalization constant and h is a filtering parameter. A limitation of this approach is the necessity of sufficient contrast in the HR scan. For regions where this is not fulfilled the upsampled anatomical structures will by limited by the lower resolution (see Fig. 2). If the squared intensity distance of the voxels in the HR image $\mathbf{z}$ is replaced by the sum of squared differences (SSD) of small images patches around these voxels, Eq. 3 can be seen as joint version of the popular non-local means filter [3]. In [7], the filter parameter h is chosen empirically and is reduced by a factor of 2 after each iteration, thus introducing a coarse-to-fine scheme of the SR reconstruction. The reconstruction process is performed iteratively by alternating steps of guided reconstruction (Eq. 3) and enforcement of the SSC (Eq. 2).

3 Robust SR using Multi-Modal Registration

In previous work [7], [8] the geometric transformation W in Eq. 1 was assumed to be known a priori. However, as shown in [7], even slight initial misalignments cause the reconstruction performance to deteriorate. The reconstructed SR volume is at best as good as an interpolated version of the LR image (without using the HR guidance). In some cases, it can also cause artifacts in the reconstructed SR image, as shown in Fig. 2 (c).

We address this problem by incorporating deformable multi-modal registration between the LR volume $\mathbf{x}$ and the HR volume $\mathbf{z}$. In [5] a point-wise multi-modal similarity metric is introduced, which can cope with the complex nature of multi-modal similarity. A multi-dimensional descriptor is computed at each location in both images, which is modality-independent and discriminative to prominent image features (such as edges and corners). We follow a similar approach as [5] and optimise a cost function consisting of the voxel-wise L2 norm of the multi-dimensional descriptors and a diffusion regularisation in a multi-resolution Gauss-Newton framework. To reduce the influence of smoothing due to interpolation, we apply the resulting deformation to the guidance image and recalculate the weights in Eq. 3 based on the transformed image at each iteration.

When initialised with SR=LR, our proposed approach consists of three alternating steps:

1. multimodal deformable registration of SR and HR scans to find W

2. reconstruction based on guided filtering using the inter-modality HR scan ($r_{\max} = 3$)

3. intensity correction to ensure subsampling consistency (SSC).

4 Experiments and Results

We demonstrate our proposed method on three different datasets. First, we use a simulated 3D MRI T1/T2 brain phantom (Brainweb [4]), which has also been used in [7] and [8] and allows us to compare our method with published results. Second, we use the Visible Human Dataset [1] consisting of 3D MRI chest scans with T1 and T2 weighting, which have been acquired post-mortem and therefore do not have any significant misalignement. For these two datasets, we artificially subsample one of the two scans to obtain a low-resolution volume. Third, we use the same chest MRI scan together with an additional CT scan of the same subject, where there exists a residual non-rigid mis-alignement between the two scans. The reconstructed SR volume $\hat{\mathbf{x}}$ is compared with the known ground truth volume $\mathbf{x}$, and the reconstruction error can be defined as mean squared error (MSE) or peak to noise ratio (PSNR) in dB: $MSE = \frac{1}{n}\sum(\hat{\mathbf{x}} - \mathbf{x})^2$ and $PSNR = 10\log_{10}\frac{(\max\hat{\mathbf{x}})^2}{MSE}$.

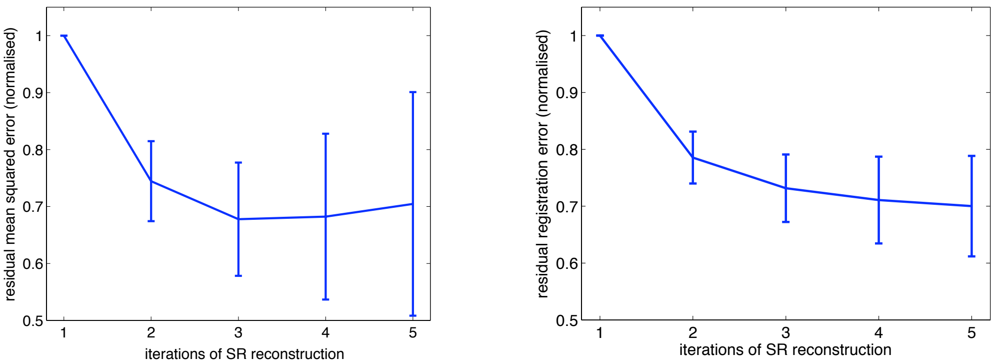

Figure 1: Residual error after each iteration of our proposed SR reconstruction method applied to the Brainweb dataset and averaged over different subsampling factors (2, 3, 4 and 6 mm). Mean squared error (MSE) is displayed on the left, the target registration error (TRE) on the right. Results are normalised by the value of the first iteration.

Table 1: Performance of reconstruction / interpolation methods (PSNR in dB) for noise-free Brainweb (BW) volumes and chest scans of the Visible Human Dataset (VHD). The BW HR guidance image (T1-weighted MRI) has isotropic voxel-size of 1.0 mm^3. The VHD HR scan (MRI-T1) has a voxel-size of 1.9x4.0x1.9 mm^3. A random subpixel shift of up to 3 mm has been applied to the HR scan before reconstruction. The LR scan is a T2-weighted MRI.

Image dataset	subsampling	NN	B-spline	**proposed**	w/o registration
Brainweb	2 mm	27.9 dB	31.5 dB	**33.7 dB**	28.2 dB
	3 mm	24.0 dB	26.0 dB	**30.5 dB**	25.0 dB
	4 mm	22.5 dB	23.7 dB	**28.1 dB**	23.6 dB
	6 mm	20.5 dB	21.2 dB	**23.9 dB**	21.4 dB
Visble Human	2 mm	31.6 dB	33.6 dB	**33.6 dB**	29.7 dB
	3 mm	28.7 dB	30.0 dB	**30.8 dB**	26.9 dB
	4 mm	27.2 dB	28.3 dB	**29.6 dB**	25.6 dB
	6 mm	25.5 dB	26.4 dB	**27.7 dB**	24.0 dB

We perform our proposed SR reconstruction on LR volumes with varying subsampling factor (2, 3, 4 and 6 mm). A random (subpixel) translation of up to 3 mm is applied for the first and second experiment to simulate initial misalignments. The resulting MSE between reconstructed SR and the ground truth volume and the residual registration error after each iteration are visualised for the first experiment in Fig. 1. To enable comparison between different subsampling factors the results are normalised by the first value of the first iteration.

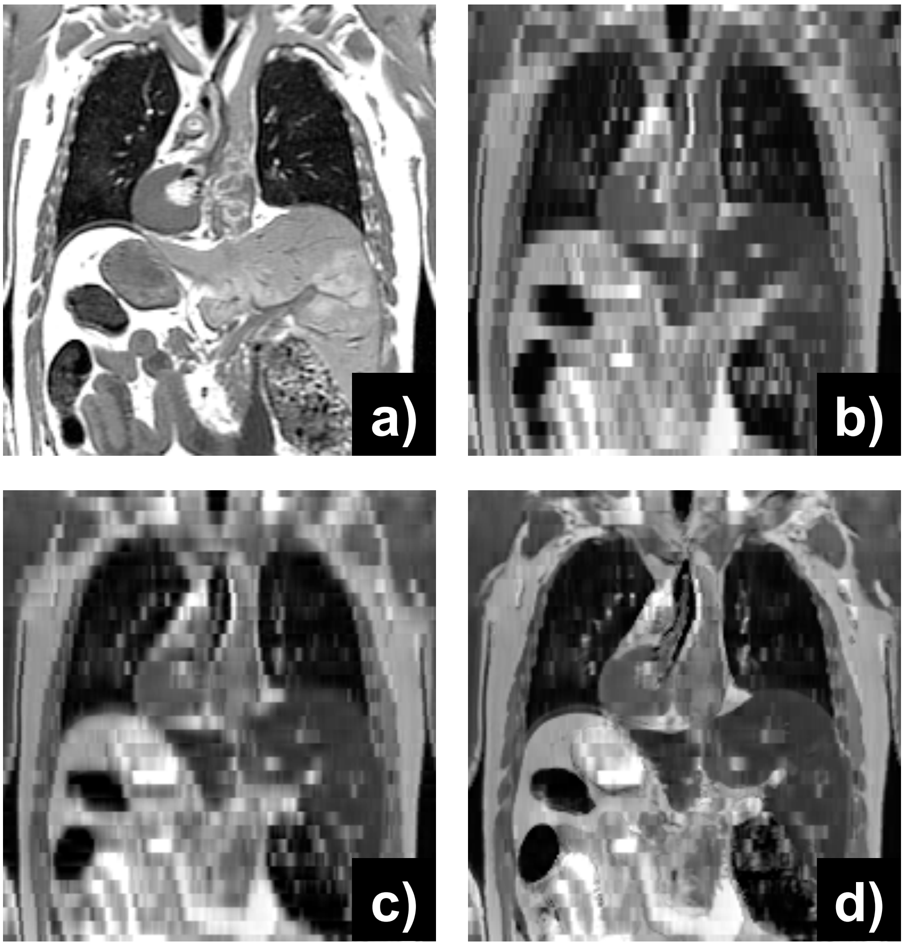

Figure 2: Coronal plane of chest MRI of VHD. (a) HR T1-weighted guidance volume. (b) LR T2-weighted volume (subsampled by a factor of 6). (c) SR reconstruction without registration. (d) SR reconstruction of our proposed method yields are visually clearly improved result. Problematic are areas, which have no corresponding anatomy in the HR scan.

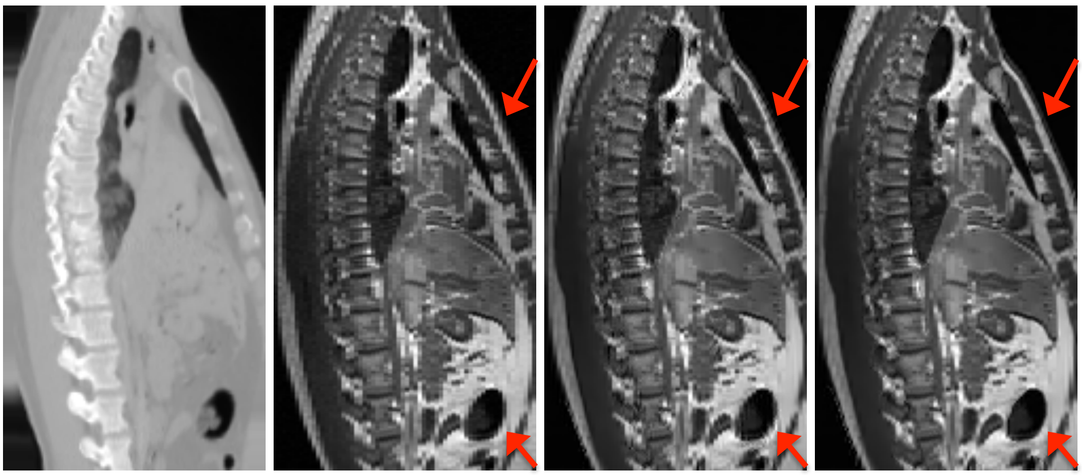

Figure 3: Sagittal plane of chest CT/MRI scan pair. From left to right: HR CT guidance volume, LR T1-weighted volume (NN-interpolation), SR reconstruction without registration, SR reconstruction of our proposed method. The red arrows indicate areas in where our proposed method achieves an improved SR reconstruction (reduced aliasing along edges).

For both measures the initial error could be further reduced by 30% on average using our proposed technique. An overview of the results for the first two experiments is given in Table 1 and an example outcome of the SR reconstruction of a chest MRI upsampled by a factor of 6 is shown in Fig. 2.

For the third experiment, no ground truth is available, therefore only visual results can be presented. The MRI scan (T1-weighted) has a resolution of 1.875x4.0x1.875 mm, the HR CT reference of 0.9x0.9x1.0 mm. We use the proposed method to construct a super-resolution MRI volume with a voxel-size of 1.875x2.0x1.875 mm. The scans are manually rigidly aligned, but there is still some minor non-rigid misalignment, due to slightly changed body positioning and geometric distortion of the scanners. Figure 3 shows an improved reconstructed volume compared to the nearest neighbour (NN) interpolation and the guided SR reconstruction without reconstruction (following the approach of [7]).

5 Conclusion

We have presented a new method to reconstruct a super-resolution (SR) volume based on a low-resolution scan and an inter-modality high-resolution guidance volume. Our technique incorporates a multi-modal registration step for a robust SR reconstruction. We demonstrate on a number of challenging clinical datasets, that the registration accuracy and the SR reconstruction can be improved by combining both methods into a unified framework. One remaining challenge are areas in the LR, which have no corresponding anatomy in the HR scan. In the future, we would like formulate our approach in a combined energy functional (registration and SR reconstruction) and apply it on a larger set of clinical CT/MRI scans.

References

[1] M.J. Ackerman. The visible human project. *Proc. of the IEEE*, 86(3):504 –511, 1998.

[2] J. Banerjee and C.V. Jawahar. Super-resolution of text images using edge-directed tangent field. In *Document Analysis Systems, 2008. DAS 2008.*, pages 76 –83, sept. 2008.

[3] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR 2005*, volume 2, pages 60–65, 2005.

[4] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transaction on Medical Imaging*, 17(3):463–468, 1998.

[5] M.P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. Gleeson, M. Brady, and J.A. Schnabel. Non-local shape descriptor: A new similarity metric for deformable multimodal registration. In *MICCAI 2011*, volume 6892 of *LNCS*, pages 541–548. 2011.

[6] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics (SIGGRAPH 2007)*, 26(3), 2007.

[7] J. Manjon, P. Coupe, A. Buades, D.L. Collins, and M. Robles. MRI superresolution using self-similarity and image priors. *Intl. J. Biomed. Imag.*, 2010:1 – 11, 2010.

[8] F. Rousseau. A non-local approach for image super-resolution using intermodality priors. *Medical Image Analysis*, 14(4):594 – 605, 2010.