SUPER-RESOLUTION RECONSTRUCTION OF CARDIAC MRI USING COUPLED DICTIONARY LEARNING

Kanwal K. Bhatia[1], Anthony N. Price[2], Wenzhe Shi[1], Jo V. Hajnal[2], Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Imperial College London, London, UK
[2]Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK

ABSTRACT

High resolution 3D cardiac MRI is difficult to achieve due to the relative speed of motion occurring during acquisition. Instead, anisotropic 2D stack volumes are typical, and improving the resolution of these is strongly motivated by both visualisation and analysis. The lack of suitable reconstruction techniques that handle non-rigid motion means that cardiac image enhancement is still often attained by simple interpolation. In this paper, we explore the use of example-based super-resolution, to enable high fidelity patch-based reconstruction, using training data that does not need to be accurately aligned with the target data. By moving to a patch scale, we are able to exploit the data redundancy present in cardiac image sequences, without the need for registration. To do this, dictionaries of high-resolution and low-resolution patches are co-trained on high-resolution sequences, in order to enforce a common relationship between high- and low-resolution patch representations. These dictionaries are then used to reconstruct from a low-resolution view of the same anatomy. We demonstrate marked improvements of the reconstruction algorithm over standard interpolation.

1. INTRODUCTION

The speed of cardiac and respiratory motion present during Magnetic Resonance Imaging (MRI) of the heart makes the acquisition of high-resolution 3D cardiac image volumes a difficult task. Instead, stacks of high resolution 2D planar slices are typically acquired, resulting in anisotropic volumes of, for example, $1.25 \times 1.25 \times 10$ mm voxels as standard. Improving the resolution of such images is an important challenge in cardiac image visualisation and analysis.

Attempts to improve resolution of images of moving structures are typically carried out either in the acquisition stage, by aiming to improve speed [1], or in the post-processing stage, based on the fusion of data from multiple acquisitions to achieve *super-resolution* [2]. Previous work in the latter approach depends on the quality of alignment of the images. Unfortunately, the highly non-rigid motion present in cardiac imaging, together with the low resolution of the acquired images makes accurate registration difficult to achieve,

setting a limit on the possible enhancement obtainable. Likewise, reconstruction techniques used for brain imaging [3], which compensate only for rigid or affine motion between acquisitions, are insufficient for cardiac image super-resolution. For this reason, upsampling of cardiac images via intensity-based interpolation schemes, such as b-spline or bicubic, are still commonly used in cardiac imaging.

We explore instead the idea of upsampling acquired data through *example-based super-resolution* using image patches. This enables the integration of information from multiple views but without the need for any non-rigid alignment or modelling of the underlying motion. Instead, we aim to exploit the data redundancy - at a patch level - across a cardiac sequence. Recent work using patch-based approaches for reconstruction has focused on brain MRI, for example to reconstruct low-resolution T2-weighted images from corresponding high-resolution T1-weighted images [4][5]. The motivation for patch-based super-resolution in cardiac imaging is even stronger than in brain imaging, due to the lack of existing acquisition techniques or post-processing algorithms that deal robustly with non-rigid motion.

The basic idea behind these approaches is that of pattern matching. Given a low-resolution (LR) image, find similar LR images in a database of LR and high-resolution (HR) image pairs. The corresponding HR pairs are then used to upsample the test image using the observed relationship between images. Since this is not possible on whole images, the methods use image patches of, for example, size $5 \times 5 \times 5$ voxels, instead. In previous works, the database of training patches consists of either other patches within the image itself or those from a HR view of the same anatomy [4][5]. Most recently, patch-based reconstruction using HR external atlases has been proposed [6]. These cases however, work on the assumption that the same relationship between the construction of HR and LR patches actually exists.

Recent developments in *sparse representations* and *dictionary learning* have begun to filter through from computer vision to the medical imaging community. In this paper, we adopt the algorithm of [7], which was developed for single image scale-up and recently applied to brain image upsampling in [8]. The key is that no relationship between the construction of HR and LR patches needs to be assumed. Instead,

such correspondence is *enforced* to exist by co-training LR and HR dictionaries, such that a LR patch has the same sparse representation with respect to the LR dictionary as its corresponding HR patch has to the HR dictionary. Using these, we are able to reconstruct HR images with up to 8 times upsampling in the slice-select direction, without any registration.

2. BACKGROUND

Standard cardiac imaging techniques typically involve the acquisition of multiple 2D slice stacks creating an anisotropic 3D volume which is HR in-plane, but LR in the through-plane direction. Given an underlying unknown HR image $\mathbf{y}_H$, the acquired LR image $\mathbf{y}_L$ can be modelled as:

$$\mathbf{y}_L = (\mathbf{y}_H * \mathbf{B}) \downarrow s + \eta \qquad (1)$$

where $\mathbf{B}$ represents a blur operator, $\downarrow s$ is a downsampling operator that decreases the resolution by a factor of s and η represents an additive noise term. Recovering the high resolution image $\mathbf{y}_H$ from $\mathbf{y}_L$ is under-determined and requires some regularisation on the nature of $\mathbf{y}_H$. In this work, we adopt the prior that small image patches of $\mathbf{y}_H$ can be sparsely reconstructed with respect to an appropriate dictionary, in the same way as $\mathbf{y}_L$, under particular circumstances.

Typically, orthogonal HR short-axis (SA), four-chamber and two-chamber long-axis (LA) views are acquired. Given a HR sequence covering the same anatomy as the LR sequence, we hypothesise that (if working with small enough image patches), there should be enough redundant information in the sequence to reconstruct a HR patch from the available data, even without information about spatial correspondences.

Fig. 1. Typical cardiac image acquisitions. L-R: HR SA slice, LA stack of short-axis slices, frame of HR LA slice sequence.

Super-resolution reconstruction (SSR) uses multiple views of an object to improve image resolution. Its application to cardiac image enhancement is attractive due to the potential to combine data acquired over time and in multiple orientations. A review of SRR in general medical imaging can be found in [9]. Techniques fall into two main categories: those based on reconstruction through the fusion of multiple views of an object [3][2], and *example-based super-resolution*, which aims to upsample LR images via knowledge of the relationship between HR and LR image features from example data [4][5]. Due to difficulties in aligning cardiac images to sufficient sub-pixel accuracy, we here adopt the latter approach. We use HR LA slices through time as training

Fig. 2. Reconstruction using co-trained dictionaries. HR and LR dictionaries are co-trained to encode corresponding patches with the same sparse representation. LR patches from a test image are then sparse coded using the LR dictionary, and the resulting sparse code applied to the HR dictionary to reconstruct a corresponding upsampled patch.
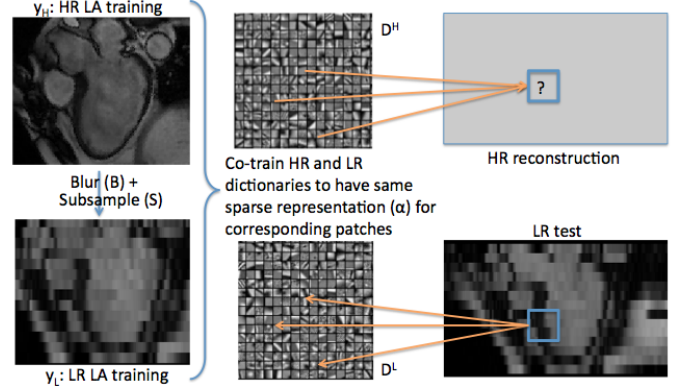
examples to reconstruct the LR orientation of a stack of SA slices.

Recent advances in computer vision have shown that signals such as image patches may be represented by a sparse combination of *atoms* from a *dictionary*. Such a dictionary can be pre-defined for general images e.g.: DCT, or learnt by training examples of similar images. A standard method of training a dictionary $\mathbf{D}$ to sparsely represent a set of signals $\mathbf{x}$ is the K-SVD method [10]. Using dictionary learning for image SRR has recently been proposed in, for example, [7]. These work on the idea that dictionaries trained from HR and corresponding LR examples can be used to reconstruct from new LR input. In the following, we describe the method of [7] and show its application to cardiac images.

3. METHODS

We aim to reconstruct a LR orientation (the through-plane view of a 2D slice stack) sequence, from frames of a HR sequence of the corresponding anatomy. To ease explanation, we assume in the following a LR SA stack sequence and single slice HR LA sequences. However, the same methods equally apply to any pair of orthogonal views that yield HR in-plane. By using orthogonal views of the same anatomy, we eliminate the need for affine registration. The method is summarised in Fig. 2.

3.1. Training

Our training data consists of frames from a HR LA sequence $Y_H = \{\mathbf{y}_H\}$, and corresponding LR sequences obtained through blurring and downsampling according to Eq. 1. By using training data that covers the whole cardiac cycle, we

can account for the variation of motion likely in the test image. A Gaussian kernel, with full width half max (FWHM) equal to the slice thickness, is used to blur in the slice-select direction only, in order to simulate the acquisition process [9]. The resulting LR images are then upsampled using bicubic interpolation back to the original size, giving $Y_L = \{\mathbf{y}_L\}$. Corresponding patches $P = \{\mathbf{p}_k^H, \mathbf{p}_k^L\}_k$ are extracted from these images: $\mathbf{p}_k = \mathbf{R}_k \mathbf{y}$, where $\mathbf{R}_k$ is an operator to extract a patch of size of size $n \times m$ from location k. These patches are used to co-train the HR and LR dictionaries.

3.1.1. Patch pairs construction

To focus the training on the relationship between LR patches and high-frequency information (edges and textures), patches are extracted from a LR image, y_L, and a difference image given by $\mathbf{y}_E = \mathbf{y}_H - \mathbf{y}_L$. For the LR patches, features containing high frequency information are extracted from y_L by convolving with 2D Gaussian / Laplacian filters: $\hat{\mathbf{p}}_k^L = f_k * y_L$. Finally, Principal Component Analysis is used to reduce the feature vector extracted: $\mathbf{p}_k^L = \mathbf{C}\hat{\mathbf{p}}_k^L$, where $\mathbf{C}$ is a projection operator that transforms $\hat{\mathbf{p}}_k^L$ to a low-dimensional subspace preserving 99.9% of its average energy. Image intensities are used for the HR patches $\mathbf{p}_k^H = \mathbf{R}_k \mathbf{y}_E$. This gives pairs of co-occurring patches at each location k, $P = \{\mathbf{p}_k^H, \mathbf{p}_k^L\}_k$.

3.1.2. Correlated dictionary learning

Correlated dictionaries ensure that a HR patch and its LR counterpart have the same sparse representations in their respective dictionaries. We also need to ensure that the LR patches can be encoded sparsely *and in the same way* for both train and test data. When reconstructing a LR test patch, we find the sparse representation of that patch in terms of the LR dictionary only. The LR dictionary $\mathbf{D}^L$ is therefore also constructed using LR patches only:

$$\mathbf{D}^L, \alpha = \min_{\mathbf{D}^L, \alpha_k} \sum_k \|\mathbf{p}_k^L - \mathbf{D}^L \alpha_k\|_2^2 \text{ subject to } \|\alpha\|_0 < \lambda \quad (2)$$

where λ denotes the desired sparsity of the reconstruction weights vector α. This standard dictionary learning equation is solved sequentially for $\mathbf{D}^L$ and α using the method of [10]. The resulting sparse code α_k for each patch k, is then used to solve for the HR dictionary by minimising:

$$\mathbf{D}^H = \min_{D_H} \sum_k \|\mathbf{p}_k^H - \mathbf{D}^H \alpha_k\|_2^2 = \min_{\mathbf{D}^H} \|\mathbf{P}^H - \mathbf{D}^H \mathbf{A}\|_F^2 \quad (3)$$

where columns of $\mathbf{P}$ are formed by the HR training patches, $\mathbf{p}_k^H$ and columns of $\mathbf{A}$ are formed by the atoms α. Denoting $\mathbf{A}^+$ as the Pseudo-Inverse of $\mathbf{A}$, the solution is given by:

$$\mathbf{D}^H = \mathbf{P}^H A^+ = \mathbf{P}^H \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \quad (4)$$

resulting in correlated dictionaries, $\mathbf{D}^H$ and $\mathbf{D}^L$.

3.2. Reconstruction

Patches from a LR test image are extracted in the same way as for the LR training images (3.1.1). The sparse code for each patch with respect to the LR dictionary, $\mathbf{D}_L$, is found by:

$$\alpha_k = \arg \min_{\alpha_k} \sum_k \|p_k^L - \mathbf{D}^L \alpha_k\|_2^2 \text{ subject to } \|\alpha_k\|_0 < \lambda \quad (5)$$

again as in [10]. Crucially, this is the same sparse coding equation as in the training phase. The reconstruction weights vector α_k for each test patch are used to approximate HR patches by $\{\tilde{\mathbf{p}}_k^H\}_k = \{\mathbf{D}^H \alpha_k\}_k$. To create a smooth overall reconstruction, overlapping patches are used, and the upsampled image given by their average reconstruction. The final HR image is given by adding the LR interpolated approximation $\tilde{\mathbf{y}}_H = \mathbf{y}_L + \left(\sum_{k \in \Omega} \mathbf{R}_k^T \mathbf{R}_k\right)^{-1} \sum_{k \in \Omega} \mathbf{R}_k^T \tilde{\mathbf{p}}_k^H$.

4. EXPERIMENTS

We demonstrate our algorithm on synthetically-generated data and use it to upsample real adult and neonatal datasets. Adult data were acquired on a 1.5T Philips Achieva scanner, using a 32 channel cardiac coil (TE/TR=1.6/3.2ms, FA=60), giving a reconstructed resolution of 1.25x1.25mm slices, 10 mm thick. SSFP acquisition with retrospective cardiac gating provides 30 frames. For the neonatal data, SSFP acquisition with retrospective ECG gating over 20 cardiac phases was used, with resulting resolution 0.5×0.5×4mm.

4.1. Results

In the simulation, we take a HR LA sequence of 30 frames, withholding one frame as a test image. This is downsampled by a factor of 8 in the theoretical slice-select direction using a Gaussian blur with FWHM=8. The remaining 29 frames are used for training using the same downsampling operations. We apply our algorithm using overlapping patches of size 16×16 pixels and dictionary size of 400 atoms. The high degree of patch redundancy allows successful reconstruction as shown in Figure 3, with substantial improvement over bicubic interpolation. Training and reconstruction took approximately 2 seconds on a 2.4GHz Intel Core i5 machine. Repeating the experiment for similar LA sequences of 15 different subjects, gave a mean mean squared error (MSE) of 2066 for bicubic interpolation compared with 856 for our patch-based method. For real MRI, we acquire a SA stack plus a HR LA sequence of an orthogonal view for a given subject. The LA sequence is downsampled and used as training data to reconstruct the corresponding non-planar LR view of the SA stack to isotropic, using an upsampling factor of 8. We present visual results for two markedly different cases: adult and neonatal. In these cases, patches of sizes of 12×24 pixels were used in both training and reconstruction. The results of two example frames from each dataset are shown in Fig. 4.

Fig. 3. Synthetically-generated example. Bicubic MSE=2683, proposed MSE=1293.

(a) Original   (b) Downsampled

(c) Bicubic   (d) Proposed
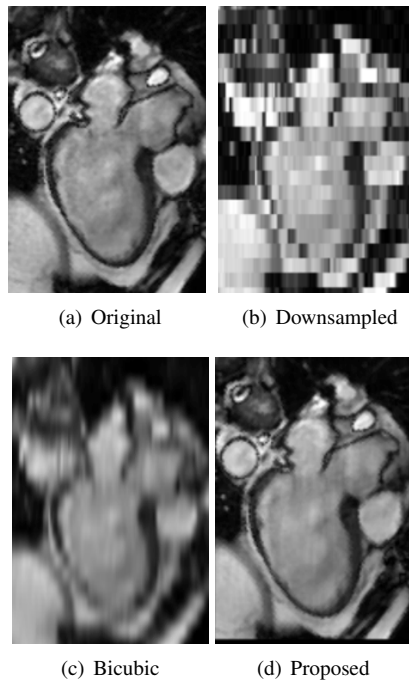
It can be clearly seen that in both cases the patch-based reconstruction algorithm produces sharper images than bicubic interpolation.

5. DISCUSSION

We have demonstrated how we can already achieve marked improvement over bicubic interpolation on real MRI data. Extensive validation of real cardiac datasets is difficult due to the lack of ground truth and is the subject of ongoing work.

6. REFERENCES

[1] J. Caballero, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity in dynamic mri," in *MICCAI (1)*, 2012, pp. 256–263.

[2] S. U. Rahman and S. Wesarg, "Combining short-axis and long-axis cardiac mr images by applying a super-resolution reconstruction algorithm," in *SPIE Med. Im.*, 2010, p. 7623.

[3] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust super-resolution volume reconstruction from slice-acquisitions: application to fetal brain mri," *IEEE Trans. Med. Imag.*, vol. 29(10), pp. 1739–1758, 2010.

[4] J. V. Manjon, P. Coupe, A. Buades, D. L. Collins, and M. Robles, "Mri superresolution using self-similarity and image priors," *Int J. Biomed. Imaging*, 2010.

Fig. 4. Example reconstruction of real MRI frames. First two rows: adult dataset. Bottom two rows: neonatal dataset.

(j) Original   (k) Bicubic   (l) Proposed

[5] F. Rousseau, "A non-local approach for image super-resolution using intermodality priors," *Medical Image Analysis*, vol. 14, pp. 594–605, 2010.

[6] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, A. Marvao, T. Dawes, D. O'Regan, and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in *MICCAI*, 2013.

[7] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *7th International Conference on Curves and Survaces*, 2012.

[8] A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain mr images using overcomplete dictionaries," *Medical Image Analysis*, vol. 17(1), pp. 113–132, 2013.

[9] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52(1), 2009.

[10] M. Aharon, M. Elad, and A. Bruckstein, "Ksvd: an aglorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54(11), 2006.