SINGLE IMAGE SUPER RESOLUTION OF 3D MRI USING LOCAL REGRESSION AND INTERMODALITY PRIORS

Jing Hu, Jiliu Zhou, Xi Wu

Department of Computer Science, Chengdu University of Information Technology, P.R. China, 610225

ABSTRACT

Clinical practice requires multiple scans with different modalities for diagnostic tasks, but each scan does not produce the image of the same resolution. Such phenomenon may influence the subsequent analysis such as registration or multimodal segmentation. Therefore, performing super-resolution (SR) on clinical images is needed. In this paper, we present a unified SR framework which takes advantages of two primary SR approaches – self-learning SR and learning-based SR. Through the self-learning SR process, we succeed to obtain a second-order approximation of the mapping functions between low and high resolution image patches, by leveraging a local regression model and multi-scale self-similarity; through the learning-based SR process, such patch relations are further refined by using the information from a reference HR image. Extensive experiments on open-access MRI images have validated the effectiveness of the proposed method. Compared to other state-of-the-art SR approaches, the proposed method provides more realistic HR images with sharp edges.

Index Terms— super-resolution, local regression, multi-scale self-similarity, intermodality priors

1. INTRODUCTION

In Magnetic Resonance Imaging (MRI), spatial resolution is limited by a number of factors such as acquisition time, short physiological phenomenon and so on. Clinically, to make a convincing diagnosis, multiple scans are always needed. Unfortunately, one of these scans is of substantially lower spatial resolution, which may exert a great influence on subsequent image analysis [1]. Therefore, increasing the resolution of MR image is of mportant to maintain the resolution level. The most intuitive method to increase the resolution is the classical interpolation method like bicubic and B-spline interpolation [2]. However, these standard interpolation approaches are not accurate enough, since the neighboring voxels, from which the intensity of an unknown voxel is estimated, may not contain the same tissue type as the unknown voxel. Hence, the interpolation methods are very likely to produce inaccurate estimation and even involve further distortion.

In recent years, the super resolution (SR) technology in MRI has emerged as an effective alternative to the interpolation method and can be broadly divided into two categories. In the first category, the SR methods are mainly implemented in the acquisition stage. Specifically, this type of methods manipulates the k-space data directly or configures the acquisition parameters to improve the spatial resolutions [3]. In the second category, the SR methods are used as postprocessing, which are mainly adapted from the classical natural image SR approaches. Specifically, two groups are maintained in this category: multi-image methods and single image methods. The multi-image methods reconstruct a high-resolution (HR) image from several sub-pixel shifted low resolution (LR) images [4] and their performance highly relies on the precision of the motion estimation [5-6]. Unfortunately, the highly non-rigid motion present in MRI, such as breathing movements during the image acquisition, together with the low resolution of the obtained images makes accurate registration difficult to achieve. The single-image methods aim at recovering a HR image from a single LR one and do not require any LR images of the same scene, thereby avoiding the difficulty of image registration. In fact, this advantage makes the single-image methods receive considerable attention in these days.

The basic idea behind these single-image SR methods is to learn the mapping between LR and HR image patches, whether these patches are from a database composed of extra images, or just from the input LR image itself. Single-image SR using extra images, or the learning-based SR, helps to recover missing high frequency information in the input LR image, which is the key issue in SR problem. Despite the effectiveness, the intrinsic drawback of the learning-based SR lies in the incompatibility between input LR image and training images. Besides that, for a reliable performance, the learning-based SR requires a huge dataset, which brings about a tremendous memory cost as well as a burdensome computation. Based on the observation that small patches tend to redundantly repeat themselves many times within the original image scale [7] as well as across different image scales, the other type of single-image SR, or the self-learning SR, uses the input image as the only source of LR-HR example patches [8], and is able to exploit more relevant examples [9], require less memory as well as less computational complexity, compared to the learning-based SR methods. However, one downside of these self-learning

Preprint submitted to 2016 IEEE International Symposium on Biomedical Imaging.
Received November 2, 2015.

methods is their incapability of recovering fine details in textural regions, since such details are actually missing in the given LR images.

In this paper, we present a unified SR framework that seamlessly takes advantages of both learning- and self-learning SR methods. Specifically, in order to map a patch in the low resolution space into the high resolution space, a robust second-order approximation of the nonlinear mapping function is estimated through the self-learning process. Moreover, considering that images with anatomical similarity and high quality are readily acquirable in medical imaging, the proposed method leverages the structural information of a HR image to guide the estimation of mapping function. Compared with the state-of-the-art methods [10], contributions of the proposed method are twofold:

1. Reliable reconstruction. Throughout the learning- and self-learning processes, both local structure information and non-local similarity are intimately related in a unified framework. With the aid of the patch redundancy from multiple image scales, we give a closed-form solution to a high-order derivative estimation of the LR-HR mapping function.

2. Fine structure preservation. A high-order derivative is beneficial to the preservation of complex details. More important, since the given HR image and the input LR image are often with different contrasts (e.g., LR T2W and HR T1W images), using high-order derivative helps to avoid the gray level discrepancy from different MRI modalities and discover the intrinsic similarity so as to achieve more accurate estimation.

The remainder of this paper is organized as follows: the next section details the proposed algorithm, experimental results and comparisons are demonstrated in Sec. 3. Finally, Sec. 4 concludes this paper.

2. METHODS

2.1. Preliminaries and notations

In this paper, matrices $\mathbf{I}$ and $\mathbf{H}$ are used to represent the input LR image and the output HR image, and matrix T denotes the reference HR image; $\mathbf{L}_n$ is the smoothed version of $\mathbf{I}$ and $\mathbf{L}_{n+1}$ is an enlarged version of $\mathbf{I}$ .The bolded lowercase $\mathbf{p}$ and $\mathbf{q}$ represent the column vectors of two $s\times s\times s$ image patches that are extracted from $\mathbf{I}$ and $\mathbf{H}$, respectively; $\mathbf{m}$ and $\mathbf{n}$ denote the column vectors of two $s\times s\times s$ image patches that are extracted from $\mathbf{L}_n$ and $\mathbf{L}_{n+1}$, respectively; $\mathbf{t}$ represents the column vector of an $s\times s\times s$ image patch that is extracted from $\mathbf{T}$. Without specific notification, we assume that $\mathbf{p}$, $\mathbf{q}$, $\mathbf{m}$, and $\mathbf{n}$ are related. That is, $\mathbf{m}$ is the most similar 3D patch to $\mathbf{n}$ among all other 3D patches in image $\mathbf{L}_n$; patches $\mathbf{m}$ ($\mathbf{n}$) and $\mathbf{p}$ ($\mathbf{q}$) have the same coordinate for the center voxel, as presented by the dashed lines in Fig. 1. Given that image $\mathbf{I}$ has more HF details than $\mathbf{L}_n$, $\{\mathbf{m},\mathbf{p}\}$ constitutes the self-exemplar LR-HR patch pairs in the proposed method. In addition, because $\mathbf{L}_{n+1}$ lacks the HF details in $\mathbf{H}$, $\{\mathbf{n},\mathbf{q}\}$ is another group of LR-HR patch pairs.

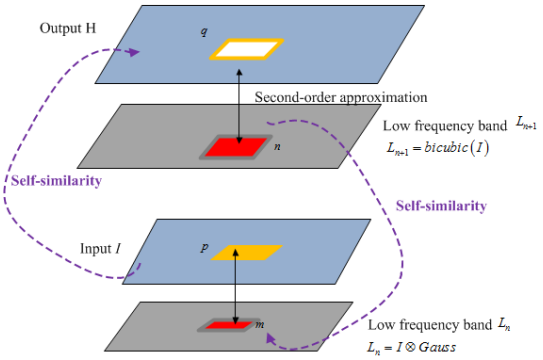

Fig. 1 Patch relations in the self-learning process

2.2. Local regression in self-learning SR

Assuming the function f associates each 3D LR-HR exemplar patch pair $\{\mathbf{n},\mathbf{q}\}$ as $\mathbf{q}=f(\mathbf{n})$, and considering that $\mathbf{m}$ and $\mathbf{n}$ are similar in terms of the geometric layout , the proposed method hence estimates the mapping function f by means of Taylor series:

$$\begin{aligned}\mathbf{q} &= f(\mathbf{n}) = f(\mathbf{m}+\mathbf{n}-\mathbf{m}) \\ &= f(\mathbf{m}) + f'(\mathbf{m})\odot(\mathbf{n}-\mathbf{m}) + \frac{1}{2}f''(\mathbf{m})\odot(\mathbf{n}-\mathbf{m})\odot(\mathbf{n}-\mathbf{m}) + \mathbf{r}_m\end{aligned} \tag{1}$$

where $\odot$ denotes the element-wise product, $f'(\cdot)$ and $f''(\cdot)$ denote the first- and second derivatives of the regression function. Moreover, by assuming the input LR image is noise-free and the mapping function follows some smoothness property, we come to $\mathbf{p}=f(\mathbf{m})$. Therefore, Eq. (1) is reformulated as:

$$\mathbf{q} = \mathbf{p} + f'(\mathbf{m})\odot(\mathbf{n}-\mathbf{m}) + \frac{1}{2}f''(\mathbf{m})\odot(\mathbf{n}-\mathbf{m})\odot(\mathbf{n}-\mathbf{m}) + \mathbf{r}_m \tag{2}$$

where $\mathbf{r}_m$ represents the residual error. According to Eq. (2), in order to obtain each HR patch $\mathbf{q}$, $f'(\mathbf{m})$ and $f''(\mathbf{m})$ should be firstly required. Su et al. have observed that a higher-order derivative helps to recover complex details [11]. However, increasing the derivative order would double the unknown number. In fact, for many existing SR approaches, only the first-order derivative is inferred. For

Preprint submitted to 2016 IEEE International Symposium on Biomedical Imaging.
Received November 2, 2015.

example, in Ref [12], the first-order derivative is assumed to be a constant, whereas in Ref. [13], such derivative is estimated from an external database. Recently, Ram et al. [14] have succeeded in estimating the first-order derivative estimation without introducing any extra training images, by translating it into a sparse coefficient computation.

2.3. Derivative estimation using multi-scale local self-similarity and intermodality priors

As illustrated in Sec. 2.2, in order to obtain a reliable first-order derivative estimation, most of the existing algorithms have so far required an external training dataset, which is highly possible to introduce false details [14]. To address this problem, in our previous work [15], we have exploited the self-similarity property so as to obtain a second-order derivative estimation. Here, we just briefly describe this estimation process and then illustrate how to refine it by introducing intermodality priors.

According to the well-known non-local patch redundancy, in image $\mathbf{L}_n$, a 3D patch $\mathbf{m}_i$ that is similar to $\mathbf{m}$ is able to be found. Therefore, in our previous work, we approximate $f(\mathbf{m}_i)$ as follows:

$$\begin{aligned}\mathbf{p}_i &= f(\mathbf{m}_i) = f(\mathbf{m}+\mathbf{m}_i-\mathbf{m}) \\ &= \mathbf{p} + f'(\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m}) + \frac{1}{2}f''(\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m}) + \boldsymbol{\xi}_i\end{aligned} \tag{3}$$

The column vector $\mathbf{p}_i$ is $\mathbf{m}_i$'s paired 3D HR patch in image $\mathbf{I}$, and the column vector $\boldsymbol{\xi}_i$ also refers to the approximation error as $\mathbf{r}_m$ does in Eq. (2). By incorporating the *J*–most similar patches $\{\mathbf{m}_i\}_{i=1}^{J}$ and their paired HR patches $\{\mathbf{p}_i\}_{i=1}^{J}$, Eq. (3) can be transformed into the following optimization problem, which can be easily solved by weighted least squares estimation:

$$\min_{f'(\mathbf{m}),f''(\mathbf{m})} \sum_{i=1}^{J} \left\| \mathbf{p}_i - \mathbf{p} - f'(\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m}) - \frac{1}{2}f''(\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m})\odot(\mathbf{m}_i-\mathbf{m}) \right\|_2^2 w_h(\mathbf{m}_i-\mathbf{m}), \tag{4}$$

where $w_h(\mathbf{m}_i-\mathbf{m})$ measures the similarity between 3D patches $\mathbf{m}$ and $\mathbf{m}_i$. It is important to note that in order to ensure a robust second-order derivative estimation, the number of equations should be above that of unknown coefficients. According to the analysis in Sec. 1, using only the input LR image cannot provide enough similar patches. To this end, we build an image pyramid by using the multi-scale image self-similarity and look for sufficient image patches in this pyramid.

Intuitively, the way of calculating $w_h(\mathbf{m}_i-\mathbf{m})$ in Eq. (4) is crucial for the final reconstruction quality. In our previous work, it only correlates with the intensity distance between two 3D patches $\mathbf{m}$ and $\mathbf{m}_i$, both of which are extracted from a LR image. However, in medical imaging, a LR image may not be simply considered as a down-sampled version of a higher resolution image. In fact, partial volume effect and geometric transformation are usually present in the LR images. Therefore, in this paper, a HR image with a different contrast is used to guide $w_h(\mathbf{m}_i-\mathbf{m})$ computation as an intermodality prior in the following way:

$$w_h(\mathbf{m}_i-\mathbf{m}) = \exp\left(-\frac{\left\|G(\mathbf{m}_i)-G(\mathbf{m})\right\|_2^2}{h^2} - \frac{\left\|G(\mathbf{t}_i)-G(\mathbf{t})\right\|_2^2}{kh^2} \right) \tag{5}$$

where $\mathbf{t}_i$ ($\mathbf{t}$) is the corresponding 3D patches of $\mathbf{m}_i$ ($\mathbf{m}$) in the guided HR image, k and h are bandwidth factors that control the decay of Gaussian function. The function $G(\cdot)$ denotes the gradient operator so as to avoid the contrast difference between LR and HR images.

3. EXPERIMENTS

In this section, the proposed method is evaluated on a publicly available MRI dataset provided by BrainWeb MRI (http://brainweb.bic.mni.mcgill.ca/brainweb/). A fixed set of parameters is used for the proposed method in all experiments. In detail, the patch size is $5\times5\times5$, the 3D searching area is a $13\times13\times3$ voxel window. The parameters h and k in Eq. (5), which are used to compute the similarity, are fixed to 10.14 and 16, respectively. The state-of-the-art algorithm NLM [10] is selected as a comparison baseline, whose source code is available at its authors' homepage. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [16] are adopted to evaluate the objective performance. A high PSNR score indicates the reconstructed HR image contains little distortion, and a SSIM value near 1 implies that the resultant HR image has a very similar structure to the ground truth.

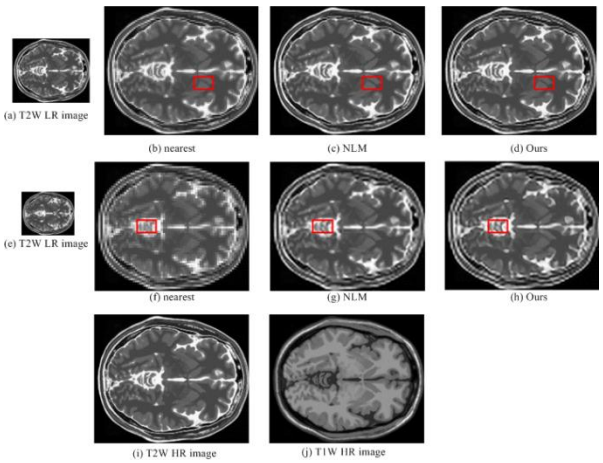

Fig. 2 Super resolution comparison. The first row: voxel resolution 2×2×2 mm^3 to 1 mm^3 voxel resolution. The second row: voxel resolution 3×3×3 mm^3 to 1 mm^3 voxel resolution. The last row shows the original HR T2W image and the reference T1W HR image.

In the BrainWeb dataset, the HR T1W and T2W volumes have 181×217×181 voxels with a resolution of 1 mm^3. To carry out simulation experiments, the T2W images of normal mode are downsampled to 2 and 3 mm isotropic resolutions and then these simulated LR data are upsampled by T1W image to 1 mm isotropic resolution under the framework of the proposed method. Fig. 2 compares the reconstruction results using different methods, and Table 1 summarizes the quantitative comparisons.

As seen in Fig. 2, the proposed method is able to provide sharper edges and more photo-realistic details than other methods. Nevertheless, it is important to point out that, PSNR metric has proven to be inconsistent with human visual perception, and it always gives a low score when the compared image is a little sharp. That is why in Table 1, the PSNR values of the proposed method are lower than that of NLM method. However, as for SSIM metric, which simulates the function of human visual system, our method obtains the highest scores.

Table 1 Comparisons on PSNR values [dB] and SSIM values (PSNR/SSIM) for different magnification factors

Factor	Nearest	NLM	Ours
2	21.74 / 0.85	26.30 / 0.94	26.18 / 0.94
3	19.44 / 0.72	21.34 / 0.82	21.19 / 0.83

4. CONCLUSION

In this paper, we present a new learning-based SR method for a single MRI image. Our model leverages the two most advanced super-resolution approaches of learning from an external HR image in other modality and learning from self-examples. By leveraging the local regression model and the multi-scale non-local patch redundancy, the proposed method establishes the relations between self LR-HR 3D patches; by using anatomical intermodality priors from a reference HR image, patch redundancy is further refined. Experimental results have confirmed the effectiveness of the proposed method on noise-free MRI images. In the future, the proposed method will be extended to address Rician noise, which is a typical issue in clinic MRI images.

.

5. ACKNOWLEDGMENTS

This work is supported by National Basic Research Program of China (973 Program) 2014CB360506 and Open Fund of Chengdu University of Information and Technology (CUIT) 762001009.

6. REFERENCES

[1] M.P. Heinrich, M. Jenkinson, M. Brady, and J.A. Schnabel, "Robust super-resolution reconstruction with multi-modal registration and guidance," *Proc. Med. Image Understand. Anal.,* pp. 81-86, 2012.

[2] R.C. Gonzalez, R.E. Woods, *Digital image processing, 2nd edition*, New Jersey: Prentice Hall, 2002.

[3] R. Shilling, T. Robbie, T. Bailloeul, K. Mewes, R. Mersereau, and M. Brummer, "A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI," *IEEE Trans. Med. Imag.,* vol. 28(5), pp.633-644, 2009.

[4] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52(1), pp. 43-63, 2008.

[5] D. Robinson and P. Milanfar, "Statistical performance analysis of superresolution," *IEEE Trans. Image Process.*, vol. 15, pp. 1413-1428, 2006.

[6] Z. Lin and H.Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 26, pp. 83–97, 2004.

[7] A. Buades, B. Coll, and J.M. Morel, "A non-local algorithm for image denoising," in *CVPR*, 2005.

[8] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *CVPR*, 2011.

[9] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Image Process.*, vol. 24(10), 2013, pp.1648-1659.

[10] J.V. Manjón, P. Coupé, A. Buades ,V. Fonov, D.L. Collins, M. Robles, "Non-local MRI upsampling," *Medical Image Analysis*, vol. 14, pp. 784-792, 2010.

[11] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multi-surface fitting," *IEEE Trans. Image Process.*, vol. 21(7), pp. 3312-3318, 2012.

[12] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph*. vol.30(2), 2011.

[13] J. Yang, Z. Lin, and S. Cohen, "Fast Image super-resolution based on in-place example regression," in *CVPR*, 2013.

[14] S. Ram, J. Rodríguez, "Single image super-resolution using dictionary-based local regression," *IEEE Southwest Symposium on Image Analysis and Interpretation*, 2014.

[15] J. Hu, Y.P. Luo, "Single-image superresolution based on local regression and nonlocal self-similarity," *Journal of Electronic Imaging*, vol. 23(3), pp. 033014-033014-14, 2014.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process*. vol. 13(4), pp. 600-612, 2004.