A Local Mutual Information Guided Denoising Technique and Its Application to Self-calibrated Partially Parallel Imaging

Weihong Guo[1], Feng Huang[2]

[1]Department of Mathematics, University of Alabama, Box 870350, Tuscaloosa, AL, USA 35487
[2] Advanced concept development, Invivo corporation, Gainesville, FL, USA 32608
[1]wguo@as.ua.edu,[2]fhuang@invivocorp.com

Abstract. The application of Partially Parallel Imaging (PPI) techniques to regular clinical Magnetic Resonance Imaging (MRI) studies has brought about the benefit of significantly faster acquisitions but at the cost of amplified and spatially variant noise, especially, for high parallel imaging acceleration rates. A Local Mutual Information (LMI) weighted Total Variation (TV) based model is proposed to remove non-evenly distributed noise while preserving image sharpness. For self-calibrated PPI, such as GeneRalized Auto-calibration Partially Parallel Acquisition (GRAPPA) and modified SENSitivity Encoding (mSENSE), a low spatial resolution high Signal to Noise Ratio (SNR) image is available besides the reconstructed high spatial resolution low SNR image. The LMI between these two images is used to detect the noise distribution and the location of edges automatically, and is then applied as guidance for denoising. To better preserve sharpness, Bregman iteration scheme is utilized to add the removed signal back to the denoised image. Entropy of the residual map is used to automatically terminate iteration without using any information of the golden standard or real noise. Results of the proposed algorithm on synthetic and in vivo MR images indicate that the proposed technique preserves image edges and suppresses noise well in the images reconstructed by GRAPPA. The comparison with some existing techniques further confirms the advantages. This algorithm can be applied to enhance the clinical applicability of self-calibrated PPI. Potentially, it can be extended to denoise general images with spatially variant noise.

1 Introduction

PPI techniques [13,16,7,14,8,2,9] are routinely used to achieve increased image resolution, decreased motion artifacts and shorter scan time to ensure patient cooperation and tolerance, as well as to meet the work flow demands of a busy clinical service. However, PPI techniques reduce acquisition time at the cost of loss in SNR and gain in residual aliasing artifacts [13]. With increase in parallel imaging acceleration rate R, the increase in noise and artifacts can be significant thereby reducing the diagnostic quality of the image.

A lot of existing denoising techniques [15,5,4,10] assume that the noise that corrupts quality of the acquired image has Gaussian distribution. However, images reconstructed by PPI have various kinds of spatially variant (non-evenly distributed) noise; and the noise distribution is decided by the geometry of the coil as well as the reconstruction scheme. Furthermore, besides noise, it is possible that there are some residual aliasing artifacts in these images. A good thing about PPI is that it provides some extra

information besides the noisy image itself. For images reconstructed by SENSE [13], the noise distribution is prescribed by geometry factor [13] that can guide the denoising [18]. For self-calibrated technique GRAPPA, a low spatial resolution high SNR image, named regulating image, is intrinsically available. The relationship between this regulating image and the noisy image is that the regulating one is reconstructed using only the fully acquired low frequency information (the central k-space), while the noisy one is reconstructed using both the low frequency and the partially acquired high frequency information. The regulating image is noise free because noise is usually of high frequency, but it is of lower resolution. The two images are perfectly registered because they are reconstructed from data created at the same time.

The purpose of this work is to develop a technique that takes advantage of the regulating image and removes spatially variant noise/artifact in images reconstructed by self-calibrated PPI, especially GRAPPA. The idea can be extended to smooth general noisy images even when there is no direct "regulating" image available. One can apply a low pass filter to generate a "regulating" image.

To protect edge information, anisotropic smoothing techniques such as the famous total variation model [15] and spatially varying diffusion filtering [11,17,3,1,6], were widely used for image denoising.

Assume the given noisy image I is related to the real clean image u by $I = u + n$ with n the noise. Under the assumption that noise n follows a Gaussian distribution with 0 mean and fixed variance, the TV model recovers u by minimizing the following energy functional.

$$A(u) = \int_{\Omega} \lambda |\nabla u| dx + \frac{1}{2} \int_{\Omega} (u - I)^2 dx \tag{1}$$

where Ω is the image domain, λ is a positive weight that is spatially invariant and is related to variance of the Gaussian noise. Constant weight λ does not work well for images with strongly non-even noise levels, especially those generated using PPI.

A substitute is to choose spatially adaptive weight λ. A straightforward one is an edge detector function that has lower values near edges but higher values in homogeneous regions. The one used in experiments for comparison is defined as follows.

$$\lambda(x) = \frac{a}{1 + b|\nabla u(x)|^2} \tag{2}$$

Edge detector functions are able to distinguish edges from homogeneous regions, but they tend to treat regions with high level of noise and artifacts that also have high gradients as edges, thereby inaccurately screen those noise and artifacts from being removed.

We propose to utilize the LMI between the regulating image and the noisy image to detect noise distribution and edge information, and to tell edges apart from highly noisey regions in particular, then use a function that is reversely proportional to LMI as weight λ and build it into a TV model. This LMI weighted TV will smooth more at noisier regions, while less near edges and smoother regions. However, if the overall weight is larger than enough, it will still smooth out fine structures and blur edges. Osher et al.'s celebrated Bregman iteration [10] comes to aid, it preserves edges by restoring edge information that might be disregarded during the smoothing process. The residual, which is the difference of the original raw image and the resulting smoothed image, often has both noise and edge information. In [10], the residual is incorporated back into

the smoothing process to restore the edge information.It is extremely important to terminate the iteration before too much noise returns to the resulting smoothed image. The original paper terminates when L_2 norm of the residual is smaller than standard deviation of the Gaussian noise. This criterion is not appropriate in practice because standard deviation is usually unknown. We propose to use entropy of the residual to terminate the iteration automatically. The proposed LMI weighted TV incorporated with Bregman iteration is named LMI-Denoiser.

The organization of the paper is as follows. In section 2, the proposed LMI-Denoiser is discussed in details. Section 3 illustrates experimental results on synthetic and in vivo data to validate the proposed model. Comparison with some other models is made in section 4. Conclusion and future work is drawn in section 5.

2 LMI-Denoiser

Mutual information and local mutual information have been widely used in image registration [12], but not in image denoising. Let I be the to-be-smoothed image, J be the regulating image. At each location x, we consider $I(x)$ and $J(x)$ as two random variables, and let $NB_r(x)$ be a window that is centered at x with side r, then use restriction of I and J on the window to approximate joint and marginal probability density function (p.d.f) of $I(x)$ and $J(x)$, the LMI between $I(x)$ and $J(x)$ is simply the mutual information between I and J at x:

$$LMI(I(x), J(x)) \triangleq \int_{NB_r(x)} \int_{NB_r(x)} p(x', y') log(\frac{p(x', y')}{p(x')p(y')}) dx' dy' \tag{3}$$

where $p(x', y')$ is the joint p.d.f. of $I(x)$ and $J(x)$, $p(x'), p(y')$ are marginal p.d.f. of $I(x)$ and $J(x)$ respectively. $LMI(I(x), J(x))$ provides a nonnegative measure of local dependence between I and J at x. It is zero only when $I(x)$ and $J(x)$ are totally independent of each other, the more they depend on each other, the higher the LMI will be. Notice that I and J are images of the same object, if there is no noise around x in I, I should dependent a lot on the noise free image J; the more noise exists at x in I, the more the dependence between I and J will be deteriorated and thus the smaller the value of LMI would be. Moreover, even though edges are blurred and fine structures are not clear in the regulating image J, it still shares strong edges, which are usually less deteriorated by noise, with I. Therefore LMI near strong edges are large. Thus, LMI detects noise distribution and strong edge location. Observing LMI is low at regions with high noise levels, high at smoother regions and near edges, a function that is reversely proportional to LMI works as an ideal denoising weight. Hence, LMI-Denoiser will be a powerful tool to guide image denoising adaptively and automatically.

The proposed LMI-Denoiser is the following.

- Initialize: $u_0 = 0$ and $v_0 = 0$.
- for k= 0, 1, 2,...: compute u_{k+1} as a minimizer of the following LMI weighted TV energy functional,

$$\int_\Omega f(LMI(I(x), J(x))|\nabla u| dx + \frac{1}{2} \int_\Omega (I + v_k - u)^2 dx \tag{4}$$

and update

$$v_{k+1} = v_k + I - u_{k+1} \tag{5}$$

where

$$f(LMI(I(x), J(x))) = \frac{a}{1 + bLMI(I(x), J(x))} \tag{6}$$

The number of iterations decides the quality of the smoothed image. We propose to utilize Shannon entropy of the residual map $I - u_k$ to terminate the iteration. Shannon entropy is a measure of uncertainty or randomness of a random variable, the more random the variable is, the higher the entropy is. At early stages, Shannon entropy of residual is low because there is more signal inside the residual, then it will increase with the iteration as more signal is removed from the residual. After it reaches the highest entropy, it will start to decrease due to return of signal in residual. The k corresponding to the global maximal entropy should provide the optimal stage. We will demonstrate that this entropy based termination criterion provides consistent results with L_2 norm and mutual information between the reference and the smoothing result.

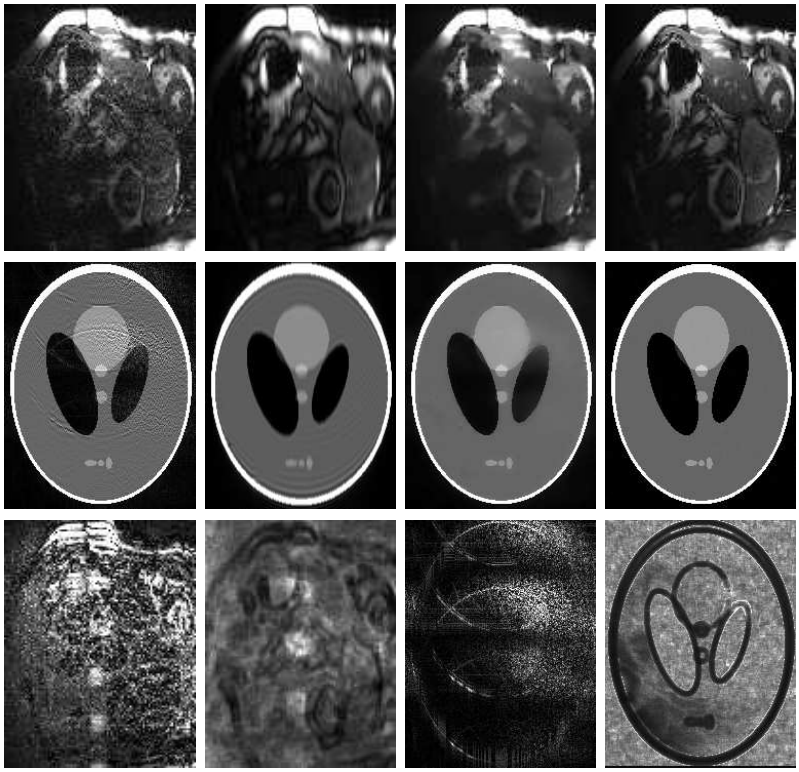

Fig. 1. Smoothing results of the cardiac and phantom data. First and second row, from left to right: the high resolution low SNR noisy image I, the regulating image J, the LMI-Denoiser smoothing result and the reference image for cardiac and phantom data respectively; third row: difference maps and $f(LMI)$ maps (as defined in 6) with $a = 10^4, b = 500$. Brighter regions correspond to higher intensity.

LMI is a p.d.f. based estimation. Theoretically, its accuracy suffers from small data size (small r), but too big r will cause wide band near edges in LMI, thus the edges of fine structures can not be detected easily. Therefore, an optimal radius is essential to the

behavior of LMI. Experimentally, we have tried $r = 4, 5, 6, 7$ for images of resolution 256×256 and found no significant difference in results, so we fix r to be 5. Moreover, we do not use normalized mutual information because the overlap between the two subimages are fixed.

There are two parameters, i.e., a, b in the model. b is used to balance the value of LMI and 1, b could be chosen such that the order of b times mean value of LMI is in the order of 10^3. b was fixed as 10^4 in the experiments. a balances the smoothing and image fidelity. Since the Bregman iteration scheme (4)-(5) can pick up signal back for over smoothed image, the choice of a can be flexible, an automatic scheme for choice of a will be provided in a separate work.

3 Experimental Results

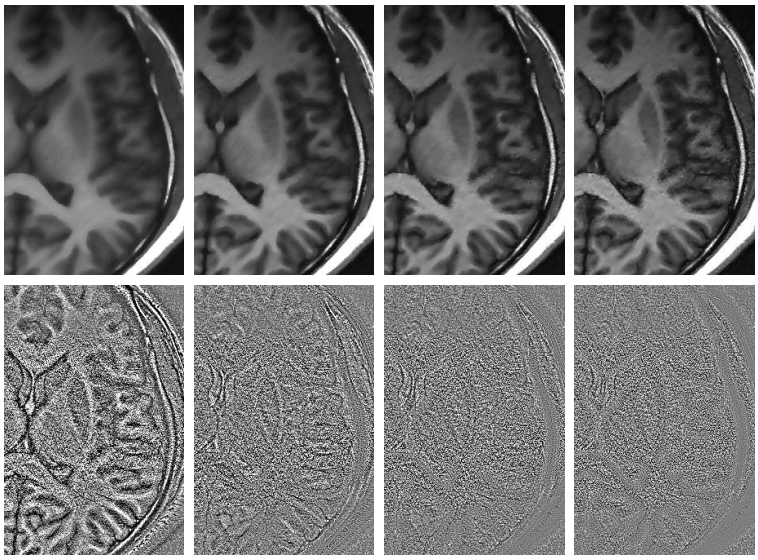

Fig. 2. Smoothing results(top) and difference maps(bottom) of four stage results of the proposed model (4)-(5). Left to right: the first to fourth stage results. $a = 100, b = 10^4$.

One set of Shepp-Logan phantom data , one set of cine cardiac data and one set of brain data are used to validate the proposed method.

Full k-space data which are not available in general are scanned in particular for these three data sets for validation purpose. The reference images, i.e. the golden standards, are reconstructed with the full k-space data. To simulate PPI, artificially down sampled data are used to reconstruct noisy image I through GRAPPA. The acceleration rate is 4, 5, and 5 for phantom, cardiac, and brain data set respectively, i.e. only $25\%, 20\%$, and 20% of the k-space data are used for reconstruction. GRAPPA with a 4×5 convolution kernel is used for image reconstruction. The convolution kernels are calculated with the auto-calibration signal (ACS) lines located at the central k-space. The number of ACS lines are 64, 24, and 56 for phantom, cardiac and brain data respectively. The ACS lines are used to generate the low spatial resolution but high SNR regulating images (J in (4)) through Fourier transform.

Fig. 1 compares the high resolution low SNR noisy image I, the regulating image J, the LMI-Denoiser smoothing result and the reference image for cardiac and phantom data. It can be seen that non-evenly distributed noise is efficiently removed and the sharpness is well preserved for both data sets. Difference between the noisy images and the reference images are shown in the third row (first image for cardiac data, third image for phantom data) to demonstrate true noise distribution. Edge information can be clearly observed from the reference image. Compare the $f(LMI)$ maps (second image for cardiac, last image for phantom) with the difference maps and reference images, it can be seen that $f(LMI)$ has lower value near evident edges and low noise level regions, and higher value near high noise level or serious artifacts region. Therefore, $f(LMI)$ detects noise distribution and strong edge information well. This explains why the smoothing results are clean and sharp.

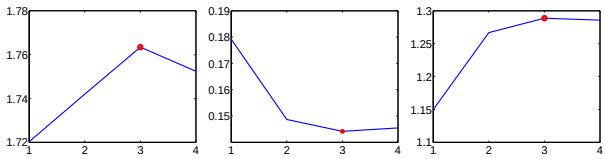

Fig. 3. Compare three termination quantities as functions of number of stages. First : entropy of the difference map between raw image and the smoothing result, second and third: relative error and mutual information of the difference map between smoothing result and the reference.

The brain data set is used to demonstrate the rationality of using entropy as a termination criterion. Fig. 2 shows the smoothing results (left) and different maps (right) after 1 to 4 stages (top to bottom). For better visibility, only partial image is demonstrated. Fig. 3 plots the entropy values of each difference map, relative errors of each smoothed image, and the mutual information values between the smoothed images and the reference images. Clearly, entropy based result matches the mutual information and relative error based results well. All of them have exactly the same tendency and show that the best result is the one at stage 3. This consistency confirms the rationality of using entropy as a termination criterion. Since entropy of difference map does not need the golden standard, which is not available in real case, entropy is the most applicable criterion in general application.

4 Discussion

This section focuses on demonstrating advantages of the proposed method over some existing TV based models. Beside the basic TV and edge detector weighted TV, we also compare with correlation weighted TV. Correlation is another natural measure of dependence between two random variables however, LMI is more general than correlation in the sense that LMI detects both linear and nonlinear dependence while correlation only detects linear dependence.

Due to range difference, parameters are chosen differently in different models, but those shown are optimal results. Fig. 4 demonstrates that basic TV, edge detector TV, and correlation weighted TV either can not remove noise sufficiently or smooth out fine structures (pointed by dashed arrows, can also be seen from the corresponding difference maps), while the proposed model is able to remove non-even noise sufficiently and

preserve sharp edges well. The result of conventional TV with small weight for smoothing has obvious residual artifacts and damaged boundary (first column); the result with large weight (second column) is artifact free, but has more seriously blurred boundary. Because the residual aliasing artifacts are serious in the noisy image, the edge detector mistakenly treats the artifacts as edges and hence smoothed the regions less. Therefore, the smoothed image (third column) has clearly residual artifacts and noise. Local correlation is another natural measure of dependence between two random variables, but it only detects linear dependence, while LMI is more general in the sense it also detects nonlinear dependence. So it is observed that the LMI based TV (last column) slightly outperforms correlation based TV (fourth column). When only linear dependence exists between I and J, and/or when computation cost is more important than accuracy, local correlation can substitute LMI.

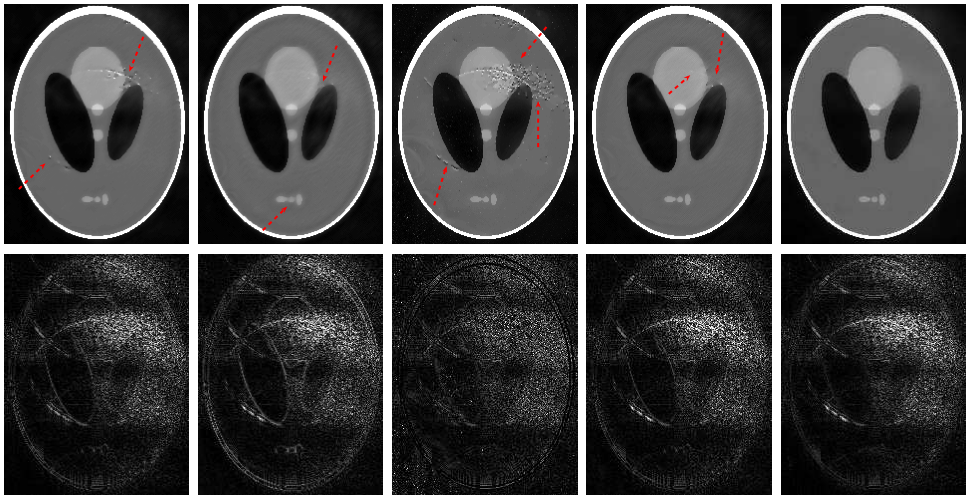

Fig. 4. Compare the proposed model with some other models. First row: smoothing results, second row: difference maps. First and second column: basic TV smoothing result with $\lambda = .3$ and $.6$ respectively; third column: edge detector TV with $a = 10^5, b = 5000$; fourth column: correlation weighted TV with $a = 1, b = .5$; last column: proposed model with $a = 400, b = 10^4$.

Quantitative comparison in sharpness and SNR between the LMI weighted TV and basic TV based model, Gradient weighted TV, correlation weighted TV are made, it shows that LMI weighted TV provides statistically significant larger sharpness and SNR.

5 Conclusion and Future work

This work proposed to use LMI between the noisy image and a regulating image to automatically detect noise distribution and strong edge information for PPI images, and then use it to remove noise and artifacts automatically and adaptively. Entropy was proved to be an applicable termination criterion for the Bregman iteration. Experimental results show the proposed method adaptively removed the noise and preserved strong edges. But LMI cannot detect small fine structures because LMI is defined based on

patch instead of point wise information. The optimal a is a function of noise/artifacts level. Hence for a given application, it is possible to decide a based on acceleration rate and the geometry of the coil, an automatic decision scheme of a will be reported in a separate work. A future work is to use the LMI between the unknown u and regulating image J to define weight. Generalization of the proposed model to any images will be reported in a separate work also.

References

1. A. Almansa, V. Caselles, G. Haro, and B. Rougé. Restoration and zoom of irregularly sampled, blurred, and noisy images by accurate total variation minimization with local constraints. *Multiscale Model. Simul.*, 5(1):235–272, 2006. 2
2. S. Banejee, S. Choudhury, E. Han, A. Brau, C. Morze, D. Vigneron, and S. Majumdar. Autocalibrating parallel imaging of in vivo trabecular bone microarchitecture at 3 tesla. *Magn. Reson. Med.*, 56:1075–1084, 2006. 1
3. M. Bertalmío, V. Caselles, B. Rougé, and A. Solé. TV based image restoration with local constraints. *J. Sci. Comput.*, 19:95–122, 2003. 2
4. A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. *IEEE Proc. Comp. Visi. Pat. Recog.*, 2:60–65, 2005. 1
5. T. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE. Trans. Imag. Proc.*, 10(2):231–241, 2001. 1
6. Y. Chen, S. Levine, and M. Rao. Functionals with $p(x)$-growth in image restoration. *SIAM Journal of App. Math.*, 66(4):1383–1406, 2006. 2
7. M. Griswold, P. Jakob, R. Heidemann, N. Mathias, V. Jellus, J. Wang, B. Kiefer, and A. Haase. A generalized approach to parallel magnetic resonance imaging: GRAPPA. *Mag. Reson. Med.*, 47:1202–1210, 2002. 1
8. F. Lin, T. Huang, N. Chen, F. Wang, S. Stufflebeam, J. Belliveau, L. Wald, and K. Kwong. Functional MRI using regularized parallel imaging acquisition. *Magn. Reson. Med.*, 54(2):343–353, 2005. 1
9. F. Lin, F. Wang, S. Ahlfors, M. Hamalainen, and J. Belliveau. Parallel MR reconstruction using variance partitioning regularization. *Magn. Reson. Med.*, 58:735–744, 2007. 1
10. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *Multiscale Model. and Simul.*, 4:460–489, 2005. 1, 2
11. P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):629–639, 1990. 2
12. J. Pluim, J. Maintz, and M. Viergever. Mutual information based registration of medical images: a survey. *IEEE Trans. Med. Img.*, 22(8):986–1004, 2003. 3
13. K. Pruessman, M. Weiger, M. Scheidegger, and P. Boesiger. Sense: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42:952–962, 1999. 1, 2
14. R. H. RM, O. Ozsarlak, P. Parizel, J. Michiels, B. Kierfer, V. Jellus, M. Mueller, F. Breuer, M. Blaimer, M. G. MA, and P. Jacob. A brief review of parallel magnetic resonance imaging. *Eur. Radiol.*, 13:2323–2337, 2003. 1
15. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 1, 2
16. D. Sodickson and C. McKenzie. A generalized approach to parallel magnetic resonance imaging. *Med. Phys.*, 28:1629:1643, 1999. 1
17. D. Strong and T. Chan. Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image. *CAM-Report 96-46 UCLA*, 1996. 2
18. S. Vijayakumar, F. Huang, J. Akao, and G. Duensing. g-factor guided denoising for sense reconstructed MR images with edge restoration. *Int. Soci. Mag. Reson. Med.*, page 2357, 2006. 2