Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2018 April ; 2018: 61–64. doi:10.1109/ISBI.2018.8363523.

THE FOURIER RADIAL ERROR SPECTRUM PLOT: A MORE NUANCED QUANTITATIVE EVALUATION OF IMAGE RECONSTRUCTION QUALITY

Tae Hyung Kim and Justin P. Haldar

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089

Abstract

In the modern biomedical image reconstruction literature, the quality of a reconstructed image is often numerically quantified using scalar error measures such as mean-squared error or the structural similarity index. While such measures provide a rough summary of image quality, they also suffer from well-known limitations. For example, a substantial amount of information is necessarily lost whenever the characteristics of a high-dimensional image are summarized by a single number. In this work, we introduce the Fourier radial Error Spectrum Plot (ESP), which provides a novel and more nuanced assessment of error by decomposing the error into its different spatial frequency components. The usefulness of ESP is illustrated in the context of MRI reconstruction from under-sampled data. In addition, we demonstrate that the extra dimension of insight provided by ESP can be used to improve the performance of existing image reconstruction techniques.

Keywords

Image quality assessment; image reconstruction; constrained MRI

1. INTRODUCTION

Quantitative image quality assessment is often a key part of the evaluation of image reconstruction methods for biomedical imaging. In the presence of a (vectorized) gold-standard reference image $\mathbf{p}_* \in \mathbb{C}^N$, it is common in the MRI literature to evaluate the quality of an image estimate $\hat{\mathbf{p}} \in \mathbb{C}^N$ using quantitative error metrics such as the normalized root mean-squared error (NRMSE) [1]

$$\text{NRMSE} = \frac{\|\mathbf{p}_* - \hat{\mathbf{p}}\|_{\ell_2}}{\|\mathbf{p}_*\|_{\ell_2}}, \quad (1)$$

the perception-inspired structural similarity index (SSIM) [2], and more specialized measures such as the high-frequency error norm (HFEN) [3] which tries to quantify the

quality of edges and other fine image features by computing the NRMSE after high-pass filtering both $\mathbf{p}_*$ and $\hat{\mathbf{p}}$.

While each of these error measures can provide useful insight into different aspects of image quality, there are many situations in which a single scalar-valued error metric is insufficient to describe the nuanced differences between different methods that can often appear in image reconstruction. A toy illustration of this is presented in Fig. 1, which shows three images that each have very different characteristics but all have the same NRMSE value.

In this work, we are inspired by the concepts of Fourier Ring Correlation (FRC) [4, 5] and Fourier Shell Correlation (FSC) [6] to develop a higher-dimensional quantitative assessment of error that provides deeper insight into the relative strengths and weaknesses of different biomedical image reconstruction approaches. FRC and FSC are techniques that were introduced in the electron microscopy literature to gauge the spatial resolution of an imaging system. These techniques operate by computing the Fourier-domain statistical correlation between two reconstructions of independent averages of the same image. These correlations are computed as a function of radius in the Fourier domain (i.e., the Fourier domain is partitioned into rings about the origin in 2D or shells about the origin in 3D, and correlation is computed separately for each ring/shell), and the results are plotted as a function of spatial frequency radius. The spatial frequency at which these correlation curves drops below a certain threshold can be used to define a statistical notion of spatial resolution [7]. However, one important limitation of FRC and FSC is that they use normalized correlation and are therefore invariant to radially-symmetric scaling errors. For example, applying a radially-symmetric Gaussian blur to an image can substantially change the quality of the image, but will not change the FRC or FSC curves.

This paper proposes the Fourier radial Error Spectrum Plot (ESP) for image quality assessment. Unlike common scalar error measures like NRMSE, SSIM, and HFEN but similar to FRC and FSC, our new ESP approach provides an extra dimension of insight that reveals frequency-dependent image quality variations. And unlike FRC and FSC, our new ESP approach is rooted in NRMSE rather than normalized correlation, and is therefore sensitive to important scaling differences that may exist between the gold standard and the reconstructed image.

For illustration of the potential insight provided by the ESP approach, ESPs corresponding to the three images from Fig. 1 are shown in Fig. 2. These ESPs reveal that even though the three images all have the same NRMSE value, they are definitely not equivalent to one another and each have their own strengths and weaknesses. In particular, the “ringing” image has the lowest relative error at low-frequencies, the “noisy” image has the lowest relative error in the mid-frequency range, and the “blurred” image has the lowest relative error at high-frequencies. This kind of insight is potentially quite valuable when choosing between different image reconstruction methods.

The following sections describe the creation of the ESP and provide illustrative applications in MRI reconstruction.

2. CREATING THE FOURIER RADIAL ERROR SPECTRUM PLOT

There are several steps to computing our proposed ESP:

1. First, we compute Fourier representations of the error image and the gold-standard image. In some MRI applications, images are reconstructed in the Fourier domain and no extra effort is required. In other cases, it may be necessary to apply a discrete Fourier transform (DFT) to the image-domain representations, e.g., $\mathbf{k}_* = \mathbf{F}\mathbf{p}_*$ and $\mathbf{k}_{\text{err}} = \mathbf{F}(\mathbf{p}_* - \hat{\mathbf{p}})$, where $\mathbf{F} \in \mathbb{C}^{M \times N}$ is an appropriate DFT operator, and $\mathbf{k}_* \in \mathbb{C}^M$ and $\mathbf{k}_{\text{err}} \in \mathbb{C}^M$ are respectively Fourier-domain representations of the gold-standard and the reconstruction error. Frequently, we will choose Fourier sampling locations such that the number of Fourier samples M is equal to the number of voxels N , though other choices are also possible.
2. Next, each of the M Fourier sampling locations is associated with a corresponding Fourier radius value, leading to sets of paired values of the form $(r_m, [\mathbf{k}_*]_m)$ and $(r_m, [\mathbf{k}_{\text{err}}]_m)$ for $m = 1, \dots, M$, where r_m is the Fourier radius corresponding to the m th Fourier sample, and $[\mathbf{k}_*]_m$ denotes the m th entry of the vector $\mathbf{k}_*$.
3. In general, multiple points in the Fourier domain will share the same Fourier radius, such that we will frequently observe multiple true signal values and multiple error values for each distinct radius value. In addition, the radius values we observe will generally be non-uniformly spaced (e.g., if $\mathbf{F}$ samples the Fourier domain on a rectilinear Cartesian grid). Moreover, the values of $[\mathbf{k}_*]_m$ and $[\mathbf{k}_{\text{err}}]_m$ will often demonstrate noise-like variations. To overcome these issues, we use smoothing splines [8, 9] to fit smooth continuous 1D functions that fit the squared error samples $(r_m, |[\mathbf{k}_{\text{err}}]_m|^2)$ and the squared true-signal samples $(r_m, |[\mathbf{k}_*]_m|^2)$ as closely as possible as a function of radius r . In particular, the smoothing spline for the squared error samples is obtained by solving

$$\hat{f}(r) = \arg \min_f \sum_{m=1}^M \left| |[\mathbf{k}_{\text{err}}]_m|^2 - f(r_m) \right|^2 + \lambda \int |f''(r)|^2 dr, \quad (2)$$

while the smoothing spline for the squared true signal samples is obtained by solving

$$\hat{y}(r) = \arg \min_y \sum_{m=1}^M \left| |[\mathbf{k}_*]_m|^2 - y(r_m) \right|^2 + \lambda \int |y''(r)|^2 dr, \quad (3)$$

where λ is a regularization parameter. It has been proven that the solutions to these two optimization problems are natural cubic splines, which enables the optimization problem to be solved using relatively easy finite-dimensional computations.

4. Finally, the Fourier radial error spectrum is obtained as:

$$\text{Error}(r) = \sqrt{\frac{\hat{f}(r)}{\hat{y}(r)}}. \quad (4)$$

Note that this ratio mirrors the same basic form as NRMSE, taking the square-root of the ratio between some measure of mean-squared error and some measure of the squared value of the true signal.

3. ILLUSTRATIVE APPLICATION TO MRI RECONSTRUCTION

To illustrate the new insight provided by ESP, we consider a head-to-head comparison between two advanced MRI reconstruction methods: SENSE parallel imaging [10] combined with total variation regularization (SENSE-TV, [11]), versus a combination of SENSE parallel imaging with low-rank modeling of local k-space neighborhoods (LORAKS) [12–14] (SENSE-LORAKS, [15]). These two reconstruction approaches are both applied to retrospectively undersampled 12-channel k-space data corresponding to the T2-weighted brain image shown in Fig. 3(a,b). This data was retrospectively undersampled using the 4.8 $\times$ -accelerated partial Fourier sampling pattern shown in Fig. 3(c).

Reconstruction results for these two reconstruction methods are shown in Fig. 4(a,b). Qualitatively, we believe that the SENSE-TV results appear somewhat blurry, while the SENSE-LORAKS results appear sharp but noisy. The conventional error metrics, which are presented in Table 1, seem to suggest that SENSE-LORAKS generally outperforms SENSE-TV with respect to NRMSE and HFEN for this dataset but that SENSE-TV has a slight advantage with respect to SSIM.

On the other hand, the ESPs shown in Fig. 5 enable a more nuanced comparison. In particular, SENSE-TV appears to have the lowest error at both the very lowest and very highest spatial frequencies, while SENSE-LORAKS has the lowest error across a broad range of middle frequencies. This observation is surprising and is not obvious from the conventional scalar error metrics, but becomes easily apparent when looking at the ESPs. These results help to confirm the potential new value offered by ESPs for comparing different image reconstruction methods.

It should also be noted that the insight provided by ESPs can be valuable for the development of improved image reconstruction methods. For example, our further analysis of SENSE-TV versus SENSE-LORAKS reconstruction suggests that the ESP characteristics for these two methods are fairly consistent across different slices in a multi-slice acquisition (results not shown due to space constraints). In particular, SENSE-TV was consistently better than SENSE-LORAKS at low-frequencies and high-frequencies, while the opposite

was consistently true across a broad range of middle frequencies. This suggests that an optimized fusion of SENSE-TV and SENSE-LORAKS (which uses SENSE-TV results at low- and high-frequencies and otherwise uses SENSE-LORAKS results) may potentially outperform either reconstruction approach individually. One such fusion image is shown in Fig. 4(c), and quantitative analysis confirms that it dominates both SENSE-TV and SENSE-LORAKS in terms of NRMSE (0.170), SSIM (0.804), and HFEN (0.212).

This example provides a potentially powerful way to use the new information provided by ESPs, yet only illustrates one such possibility. We expect that the research community will be able to find many other potential uses for this new quantitative error analysis tool.

4. CONCLUSION

This work proposed a novel error analysis tool for image quality assessment. Rather than the conventional approach in which error characteristics are summarized by a single scalar value, the ESP represents the errors as a spectrum that varies as a function of spatial-frequency. Although there is no universally best error metric (i.e., assessment of image quality is always subjective and/or context dependent), we believe that the ESP enables more nuanced insights into general image quality than are available from conventional measures. The proposed approach is expected to be useful in many different image reconstruction scenarios, and as our results demonstrate, it also enables new reconstruction approaches that have the potential to outperform conventional approaches by taking optimal account of the unique strengths and weaknesses of existing image reconstruction approaches. It should also be noted that while we only illustrated ESPs in the context of MRI reconstruction, we believe that the concept will be equally advantageous across a broad range of other biomedical imaging modalities.

Acknowledgments

This work was supported in part by research grants NSF CCF-1350563, NIH R21 EB022951, NIH R01 NS074980, and NIH R01 NS089212.

5. REFERENCES

- [1]. Wang Z and Bovik AC, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, pp. 98–117, 2009.
- [2]. Wang Z, Bovik A, Sheikh H, and Simoncelli E, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004. [PubMed: 15376593]
- [3]. Ravishanker S and Bresler Y, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, pp. 1028–1041, 2011.
- [4]. Saxton WO and Baumeister W, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *J. Microsc.*, vol. 127, pp. 127–138, 1982. [PubMed: 7120365]
- [5]. van Heel M, Keegstra W, Schutter W, and van Bruggen EFJ, "The structure and function of invertebrate respiratory proteins," in *EMBO Workshop*, 1982, pp. 69–73.
- [6]. Harauz G and van Heel M, "Exact filters for general geometry three dimensional reconstruction," *Optik*, vol. 73, pp. 146–158, 1986.
- [7]. van Heel M and Schatz M, "Fourier shell correlation threshold criteria," *J. Struct. Biol.*, vol. 151, pp. 250–262, 2005. [PubMed: 16125414]

- [8]. Reinsch C, "Smoothing by spline functions," *Numerische Mathematik*, vol. 10, pp. 177–183, 1967.
- [9]. Craven P and Wahba G, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 31, pp. 377–403, 1979.
- [10]. Pruessmann KP, Weiger M, Scheidegger MB, and Boesiger P, "SENSE: sensitivity encoding for fast MRI," *Magn. Reson. Med*, vol. 42, pp. 952–962, 1999. [PubMed: 10542355]
- [11]. Block KT, Uecker M, and Frahm J, "Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint." *Magn. Reson. Med*, vol. 57, pp. 1086–1098, 2007. [PubMed: 17534903]
- [12]. Haldar JP, "Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI," *IEEE Trans. Med. Imag*, vol. 33, pp. 668–681, 2014.
- [13]. Haldar JP, "Low-rank modeling of local k-space neighborhoods: from phase and support constraints to structured sparsity," *Proc. SPIE*, vol. 9597, p. 959710, 2015.
- [14]. Haldar JP and Zhuo J, "P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data," *Magn. Reson. Med*, vol. 75, pp. 1499–1514, 2016. [PubMed: 25952136]
- [15]. Kim TH, Setsompop K, and Haldar JP, "LORAKS makes better SENSE: Phase-constrained partial fourier sense reconstruction without phase calibration," *Magn. Reson. Med*, vol. 77, pp. 1021–1035, 2017. [PubMed: 27037836]

Fig. 1.

A gold-standard reference image (obtained from <http://mr.usc.edu/download/data/>) along with three versions of this image that each present very different kinds of errors, yet all share the same NRMSE value of 0.231 with respect to the gold standard.

Fig. 2. Proposed ESPs corresponding to the “blurred,” “noisy,” and “ringing” images from Fig. 1.

Fig. 3.

(a) Magnitude and (b) phase images of the gold-standard T2-weighted reference image. (c) 4.8 $\times$ -accelerated partial Fourier k-space sampling mask used for retrospective undersampling.

Fig. 4.

(a) SENSE-TV and (b) SENSE-LORAKS reconstruction results from undersampled data.
(c) Fusion of SENSE-TV and SENSE-LORAKS reconstruction results.

Fig. 5.
 ESPs for the SENSE-TV, SENSE-LORAKS, and their fusion shown in Fig. 4.

Table 1.

Conventional Error Metrics for SENSE-TV and SENSE-LORAKS.

	SENSE-TV	SENSE-LORAKS
NRMSE (lower is better)	0.204	0.178
SSIM (higher is better)	0.802	0.775
HFEN (lower is better)	0.330	0.216