Applying compressed sensing and a temporal basis to magnetic resonance imaging

Phil Bones, Bahareh Vafadar, and Bing Wu†
Computational Imaging Group
Dept. Electrical & Computer Engineering, University of Canterbury, Christchurch, N.Z.
† GE Medical, Beijing, China.
phil.bones@canterbury.ac.nz

***Abstract*—The application of compressed sensing (CS) in accelerating magnetic resonance imaging (MRI) is reviewed and an extension suitable for magnetic resonance angiography (MRA) is presented. The sparsity inherent in typical MR images can be exploited in order to replace regular full sampling of measurement space with sparse sampling. The authors recently proposed the enhancement of image sparsity by data ordering with the algorithm PECS. An extension to PECS is presented which projects reconstructed MRA image data onto a specially designed temporal basis. Preliminary results are presented for a normal subject.**

***Index Terms*—magnetic resonance angiography; compressed sensing; temporal basis functions.**

I. Introduction

The significant time necessary to record each resonance echo from the volume being imaged in magnetic resonance imaging (MRI) has led to much effort to develop methods which take fewer measurements. Faster methods mean less time for the patient in the scanner, increased efficiency in the use of expensive scanning facilities, improved temporal resolution in studies involving moving organs or flows, and less artifacts from patient motion. Images like those of the human body possess the property of *sparsity*, i.e., the property that in some transformed space they can be represented much more compactly than in image space. The technique of compressed sensing, which aims to exploit sparsity, has therefore been adapted for use in MRI [1]. This, coupled with the use of multiple receiving coils (parallel MRI) and the use of various forms of prior knowledge (e.g., support constraints in space and smoothness), has resulted in significantly faster image acquisitions with only a modest penalty in the computational effort required for reconstruction. We briefly introduce the technique of compressed sensing and its use in MRI. We then review the use of data ordering as a method of further exploiting sparsity and the algorithm 'PECS', which we have recently presented [2]. Sparsity also exists in the temporal domain when voxel intensities over time reflect the quantity of a contrast agent (gadolinium) within blood vessels. In the following we present our use of the Kahunen-Loeve Transform to generate a temporal basis and thereby to improve the reconstruction with PECS of contrast enhanced magnetic resonance angiography (CE-MRA) image sequences.

II. Compressed sensing

The conventional wisdom in signal processing is that the sampling rate for any signal must be at twice the maximum frequency present in the signal. However in the work performed in recent years related to signal compression, it has become obvious that the total amount of information which is needed to represent a signal or image to high accuracy is in many cases much less than that implied by the sampling theorem. This is nowhere more apparent than in the modern digital camera where quite acceptable images can be stored and recreated from a small fraction of the data volume that was associated with the original image sampling, a direct consequence of the property of sparsity. The technique of compressed sensing (also known as 'compressive sensing' and henceforth abbreviated to 'CS') was introduced to exploit image sparsity [3], [4]. Consider a 2D image with N pixels represented by the vector $\mathbf{x}$ and suppose that it can be accurately represented by $K << N$ data values under the linear transformation $\mathbf{y} = \Phi\mathbf{x}$. Rather than to measure the N pixel values and then to perform the transformation, we seek to make just M measurements $\mathbf{m}$, where $K \leq M << N$, and estimate the transformed version of the image. Thus $\mathbf{m} = \Psi\mathbf{y}$, where Ψ is a measurement matrix of dimension $M \times K$. While this might be of little direct benefit in the case cited above of a modern digital camera, for which the design of the sensor is most straightforwardly implemented as a regular 2D array of individual pixel detectors, there are many other applications, notably including MRI, for which making fewer measurements does offer an advantage.

An illustration of how CS could work in optical imaging is shown in Fig. 1. Suppose that the well known cameraman image is being acquired with a conventional digital camera in Fig. 1(a) and a full set of pixels are being recorded and stored. Applying a linear transform, such as the discrete wavelet transform (DWT), to the image allows many of the DWT coefficients to be set to zero, resulting in a compressed form of storage. The compressed data set can be used to reconstruct a good likeness of the original image. The alternative CS approach is illustrated in Fig. 1(b): by some process many fewer samples are made of the original scene and the non-zero coefficients of the compressed image are directly estimated

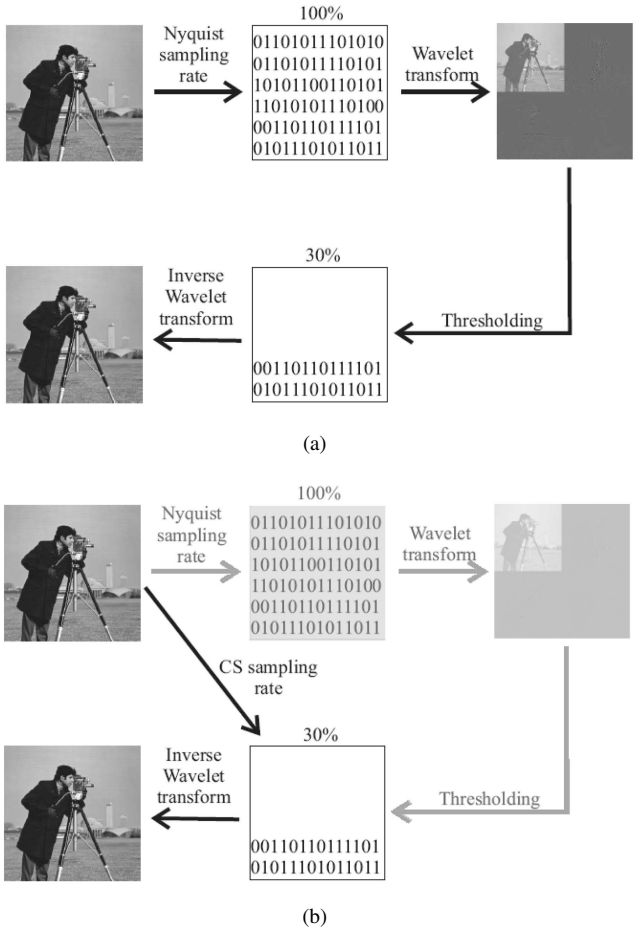


Fig. 1. Comparing (a) conventional image sensing and compression to (b) compressed sensing. In (a) the image is sampled at the Nyquist sampling rate, then all of the smallest coefficients in the image's wavelet transform are discarded. In (b) the significant wavelet coefficients are directly estimated from fewer samples of the image.

with CS. Thus the waste associated with performing full data measurement followed by compression is avoided in CS.

III. Applying compressed sensing to MRI

MRI represents the entire set of methods which apply the principles first developed for chemistry as nuclear magnetic resonance in such a way that the spatial variation of a property within an object is observed. Because of the relationship between the resonance frequency of certain atoms and magnetic field strength, virtually all MRI methods make measurements in spatial frequency space, or 'k-space' as the MRI community refers to it. The use of MRI is restricted in one specific way: the physical processes involved in the excitation and reception of MR signals are inherently quite slow. Thus the time taken to invoke a specific sequence and to measure the signals that are generated is measured in a matter of milliseconds per k-space sample. Note that this has nothing to do with the electronics associated with the scanner - faster hardware does not solve the problem. To take a complete set of measurements for, say, a 3-D imaging study of the brain may take many minutes; some acquisition sequences require periods approaching one hour. A key method to speed up acquisition has been the introduction of multiple sensing coils and corresponding algorithms to exploit the extra information, such as SENSE, for performing what is now known as 'parallel MRI' [5]. Further strenuous efforts are being made to achieve further acceleration through signal processing, including with CS.

The formal introduction of CS into MRI methods was made by Lustig, Donoho and Pauly in 2007 [1]. Their key contribution was the explicit use of a different transform domain for appropriate application of the $\mathcal{L}_1$ norm. The authors identified the use of discrete wavelet transform (DWT) and discrete cosine transform (DCT) as suitable transform bases for application in MR images, as evidenced by their sparse representation under DWT and DCT. A reconstruction framework was given, which converts the CS formulation into a convex optimization problem and hence allows for computational efficiency [1]. The authors also spelt out that a key requirement in data measurement for successful compressed sensing recovery is to achieve incoherent aliasing.

IV. Increased sparsity by data ordering

A quite distinctly different approach to increasing sparsity has recently been proposed. In 2008 Adluru and DiBella [6] and Wu et al [7] independently proposed performing a sorting operation on the signal or image as part of the reconstruction process. The principle is presented in Fig. 2 for a 2D axial brain image. In Fig. 2(a) the situation is shown whereby the image is transformed by the 2D DCT and then a compression occurs by setting all coefficients less than a given threshold to zero. The resulting reconstruction is similar to the original, but noticeably smoother due to the loss of some small amplitude high frequency components. In Fig. 2(b), the image pixels are sorted from largest amplitude in the lower right to highest amplitude in the upper left to make the resulting function monotonic and the mapping required to do this is retained (denoted 'R'). The same transformation and recovery operation after thresholding as in (a) is performed and a re-sorting (denoted 'R^{-1}') is performed. Because the compression retains much of the shape of the image after sorting, the result has much higher fidelity than in Fig. 2(a). We argue that many fewer coefficients need to be retained in the DCT of the sorted image than in the original, hence the more successful reconstruction.

Clearly the process depicted in Fig. 2 requires knowledge of the original signal in order to derive R. The practical utility of what has been demonstrated is likely therefore to be questioned. However we have demonstrated that several methods to derive an approximate R are possible and that they lead to useful and practical algorithms for MR image recovery. One such method is now described in the next section.

V. Prior estimate-based compressed sensing

We recently introduced the method 'prior estimate-based compressed sensing' (PECS) and demonstrated its use in both brain imaging, i.e., imaging of a static structure, and in contrast-enhanced angiography, i.e., dynamic imaging, as

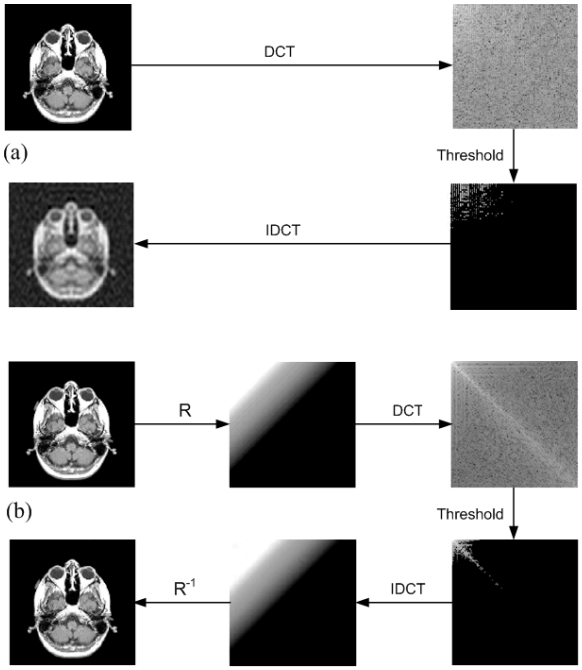


Fig. 2. Illustration of how a data ordering can achieve a higher sparsity for a 2D image. In (a) the signal is compressed by retaining only those DCT coefficients with amplitudes higher than a threshold. In (b) the image pixels are sorted to generate a monotonic function and then the same recovery operation is performed before a final resorting. Because the sorted data in (b) is more sparse, the recovery is of higher quality.

part of a pilot study on normal volunteers [2], [8], [9]. We briefly describe the method and present preliminary results for improving the reconstruction of sequences of images in CE-MRA.

The success of CS is determined by the sparsity of the underlying signal. In our experience to date, the sparsest representation of a typical MRI anatomical image is obtained by ordering the set of pixel (or voxel) amplitudes as described in Section IV. Assume that a set of under-sampled k-space data have been collected with a particular MRI sequence with the purpose of forming a high resolution image. In addition, assume a prior estimate is available of the image, for example a low resolution version of the image. In PECS the prior estimate of the image is first used to derive a data ordering, R; CS is then used to recover an image from the measured k-space data, incorporating the approximate ordering (to promote sparsity) and a total variation (TV) minimization (to promote piecewise smoothness in the resulting image) [2], [8]. When R is applied to the high resolution data, the result is a highly noisy signal which only approximates in form to a monotonic function. Under the transform the largest coefficients retain the form, while the errors tend to generate a noise-like spectrum with low amplitudes spread across many coefficients. After applying a threshold, only the significant coefficients are non-zero and these retain the form of the true ordering. We argue that prior knowledge about the signal is thereby introduced by the application of the approximate data ordering [2].

It remains to explain the formation of a prior estimate. In the case of using PECS to reconstruct CE-MRA image sequences, we use an image formed by combining all of the k-space data, i.e. a static image reconstructed by CS. This image does not retain any of the temporal variation, but it does approximate closely to the anatomy of the subject.

VI. Forming a temporal basis

In CE-MRA a bolus of gadolinium-containing compound is injected into a peripheral vein. The injected material diffuses within the bloodstream as it is propelled to the right side of the heart, through the lungs, through the left side of the heart and then finally to the site of interest. The diffusion within the moving bloodstream is such that the concentration of contrast material, observed at some downstream point, rises from zero to a peak and then decays somewhat more slowly than it rose. If monitoring continues, it may be possible to see a later recirculation phase, but attention is normally directed to the first pass of the contrast material. A number of authors have analysed the shape of the contrast-versus-time curve and concluded that it approximates to a gamma variate function [10], [11] (neglecting the recirculation, if present). In its most common form the gamma variate function can be expressed as

$$y(t) = A(t-t_0)^{\alpha}\ \exp\left(-(t-t_0)/\beta\right),\ \ t \geq t_0, \qquad (1)$$

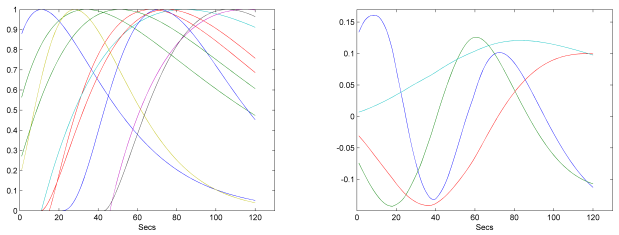
where t_0 is the point at which the function commences and α and β are real parameters. However Madsen [11] showed how it can be expressed in a more convenient form:

$$y(\tau) = y_{max}\ \tau^{\alpha}\ \exp\left(\alpha(1-\tau)\right), \qquad (2)$$

where $\tau = (t-t_0)/(t_{max}-t_0)$, t_{max} is the time at which the function peaks and y_{max} is the amplitude at that time. The shape of the gamma variate function in the form of Eq. 2 is controlled by the single parameter, α.

Within a sequence of MRA images, the expectation is that different voxels will have distinctly different temporal variation throughout the sequence. Firstly, those associated with areas free of significant bloodflow will have little temporal variation other than the inevitable signal noise and the artifacts likely to be caused if the sampling density in k-space is below the Nyquist limit. Those associated with arteries will exhibit a relatively early and rapidly rising contrast pulse. Those associated with the venous drainage will exhibit a relatively later and more slowly rising pulse with lower peak amplitude than for the arterial voxels. The time of onset will depend on the distance from the point of observation to the injection site and the timing of the injection relative to the start of the data acquisition. While the gamma variate function provides a very useful model for the expected intensity variation in those parts of the image where blood vessels lie and therefore through which blood and contrast flows, the function has several parameters controlling the temporal variation, i.e., t_0, t_{max} and α, as presented above in Eq. 2. Our approach is to project the temporal variation within each voxel onto a set of basis functions, a linear combination of which can

accurately represent the range of responses expected within the image. We have employed the Karhunen-Loeve transform (KLT) which is known to provide an optimal basis under certain conditions [12].



(a) (b)

Fig. 3. Illustration of the formation of a set of temporal basis functions: (a) the first 10 out of a set of 100 unit amplitude gamma variate functions with randomly assigned parameters, and (b) the first 4 basis functions generated by KLT from the training set.

A total of 100 prototype functions were generated, the first 10 of which are shown in Fig. 3(a). In each case the peak amplitude was fixed at unity while the three other parameters in Eq. 2 were randomly chosen with a uniform distribution from within the following ranges: t_0 -25 to +45 seconds; t_{max} +25 to +60 seconds; and α 0.8 to 3.0. These ranges were chosen so that the 100 gamma variate functions represented a reasonable approximation to the expected shape of intensity-versus-time responses in the various parts of the field-of-view (FOV) where blood vessels lie. The KLT was then applied. Fig. 3(b) shows the first four basis functions generated by the process. Experimentation indicated that a total of 4 temporal basis functions was sufficient to fit the set of generating functions within a small mean-square error, so the results presented in Section VIII are for 4 basis functions. While recomputing the basis functions for a different set of gamma variate exemplars with the same parameter ranges will generate a different basis, experience shows that the functions are very similar in nature each time. Likewise, making small changes to the ranges of the gamma-variate parameters was found to have relatively little effect on the basis functions formed by KLT, suggesting that the process is reasonably robust.

VII. In-vivo measurement

A study was performed on the knee region of a human volunteer to investigate the performance of the proposed method. A 1.5T GE scanner equipped with an 8-channel receiver channel coil, in which the individual receiver channels are symmetrically placed around the cylindrical coil, was used. A T1-weighted 3D spoiled gradient recalled (SPGR) sequence was used, and the two phase encoding directions were set to superior-inferior (SI) and anterior-posterior (AP).

Ethical approval was obtained from the Upper South A Regional Ethics Committee (Ministry of Health, NZ) to make a single injection of gadolinium contrast into the volunteer. The following scan parameters were used: $TR = 4.8$ ms, $TE = 1.8$ ms, Flip angle$= 45^o$. A matrix size of $196\times128\times48$ was used to obtain a spatial resolution of $1 \times 1 \times 1$ mm. 20 ml of Gd-BOPTA (Multihance) was injected as a bolus.

Acquisitions were defined by a sampling mask which sampled 25% of k-space and 5 repetitions of the partial k-space data sets were obtained. Samples were randomly located, but with a higher probability of being close to the centre of k-space. Each repetition took 30 s to complete. The first repetition of the data acquisition sequence was timed to complete before the contrast arrived into the region being imaged; that data was used for estimating the coil sensitivity profile and was subtracted from the following repetitions to help eliminate the stationary background. A sliding window, of width T s, on the data was employed to generate a sequence of reconstructed frames. Thus T = 7.5 s represented an acceleration factor (AF) of 16, since only one sixteenth of the k-space volume was sampled for each frame.

VIII. Results

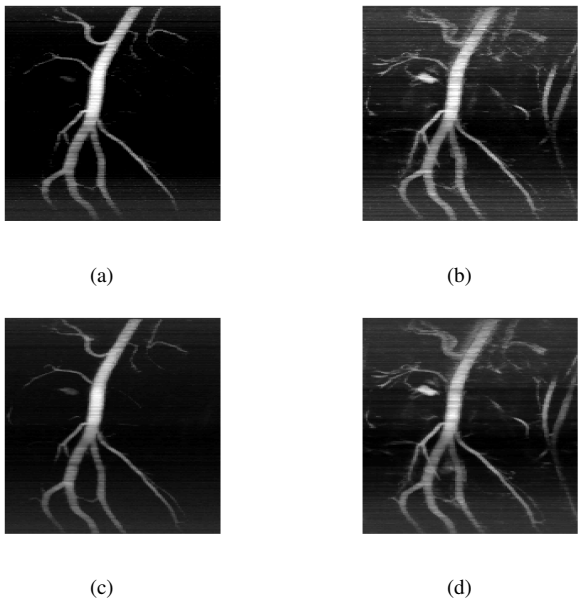
The images shown in this section are maximal image projections (MIP) viewed from the front of the subject's knee. Fig. 4 (a) and (b) show two frames from a sequence reconstructed by PECS with T = 7.5 s. One frame was generated every 1.6 s. The reconstructions show noticeable artefacts caused by the difficulties of reconstructing from such a small amount of data. Corresponding reconstructions with a temporal basis applied are shown in Fig. 4 (c) and (d), respectively. Only the first 4 temporal basis functions were used. In both cases the use of the temporal basis has had a significant effect in suppressing the background artifacts and significantly improved the clarity of the vessels within the arterial tree. The venous structure is not visible, since the frames shown are relatively early in the sequence. Note that the general quality is reasonable and there is a clear progression from the earlier to the later frame.

Fig. 5 (a) and (b) again shows 2 selected frames from a reconstructed sequence of MIP images, but this time T = 5 s., so even less k-space data was used for each frame. The frames selected are again (a) shortly after the appearance of the contrast agent and (b) near the point of peak concentration of the contrast agent. Corresponding reconstructions after projecting the data onto the temporal basis are shown in Fig. 5 (c) and (d), respectively.

In Fig. 6 plots of the standard deviation (SD) of a group of voxels within the FOV, but outside of any blood vessels, are shown. The voxels concerned were from a 11×11 voxel block in the centre slice of the FOV. The plots shown are over time, both before and after projecting the data onto the temporal basis and the sequence used is for T = 7.5 s. A significantly lower value of SD is observed after the temporal basis is applied.

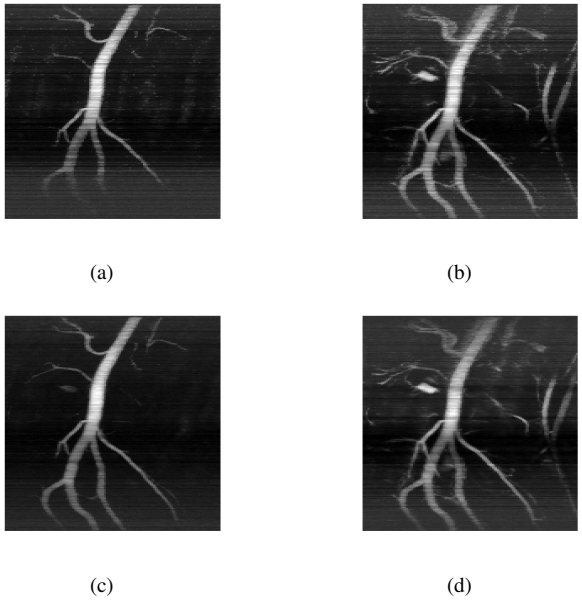
IX. Discussion

We have presented some preliminary results for projecting reconstructed CE-MRA sequences onto a temporal basis designed to match the expected dynamics of the contrast material. While imposing constraints on temporal variation



(a) (b) (c) (d)

Fig. 4. Frames selected from the sequence reconstructed by PECS with T = 7.5 s: (a) 14th, and (b) 48th frame. (c) - (d) the corresponding frames for the sequence projected onto 4 temporal basis functions generated by KLT.



(a) (b) (c) (d)

Fig. 5. Frames selected from the sequence reconstructed by PECS with T = 5 s: (a) 9th, and (b) 32nd frame. (c) - (d) the corresponding frames for the sequence projected onto 4 temporal basis functions generated by KLT.

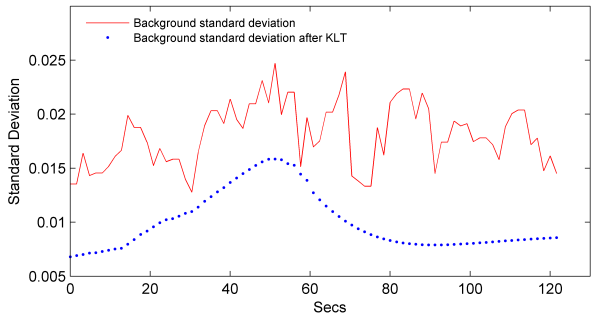


Fig. 6. Plots comparing the standard deviations for a small region (121 voxels) free of vessels with T = 7.5 s.

is not new, it has in the past primarily been applied to cardiac imaging [13], [14], where the pseudo-periodic nature of the heart activity can be exploited. As far as we are aware, ours is the first use of a temporal basis generated by KLT from a set of gamma variate prototype functions. The method has the potential to significantly improve the quality of the angiographic sequence, by suppressing frame-to-frame variation which becomes very noticeable when the sequence is viewed as a real-time movie loop.

Our method incorporates a data ordering step in the recovery of MR images by compressed sensing. There remains considerable scope for putting this non-linear processing on a firm theoretical footing. Candès and others have provided such rigour to the basic CS recovery of certain classes of image [3], but no such attention has to our knowledge been directed at the data ordering. Our conjecture is that prior knowledge of the image can be usefully incorporated in the CS process through data ordering. Experience also shows that the aliasing artefacts which are an inevitable consequence of taking fewer k-space samples are converted into distributed errors more akin to random noise; while any error is undesirable, such random noise is often less likely to obscure genuine image features.

We have demonstrated very high acceleration factors for the generation of CE-MRA image sequences. In Fig. 5, for example, each frame in the sequence is generated from only one twenty fourth of the samples needed for full Nyquist sampling of the three dimensional k-space. Such high accelerations make dynamic studies of the flow of contrast agent practical. In addition, the nature of the sampling makes retrospective choice of the acceleration practical. Thus, image quality (SNR) can be traded off against temporal resolution.

X. Conclusion

In CE-MRA the sparsity associated with the relatively smooth change in contrast agent concentration can be used to improve the quality of reconstructed image sequences. A method which projects the voxel data onto a temporal basis formed by the Kahunen-Loeve Transformation of prototype gamma variate functions has been demonstrated on a study using a human volunteer.

ACKNOWLEDGMENT

The authors would like to thank the National Heart Foundation of New Zealand for supporting this research with a project grant.

REFERENCES

[1] Lustig, M., Donoho, D., and Pauly, J.M. "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magn. Res. Med., vol. 58, pp. 1182-1195, 2007.

[2] Wu, B., Millane, R.P., Watts, R., and Bones, P.J. "Prior estimate-based compressed sensing in parallel MRI," Magn. Res. Med., vol. 65, pp. 83-95, 2011.

[3] Candes, E., Romberg, J., and Tao, T. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. on Inf. Theory, vol 52, pp. 489-509, 2006.

[4] Baraniuk, B. "Compressive sensing," IEEE Signal Proc. Mag., vol. 24, pp. 118-121, 2007.

[5] Pruessmann, K.P., Weiger, M., Scheidegger, M.B., and Boesiger, P. "SENSE: Sensitivity encoding for fast MRI," Magn. Res. Med., vol. 42, pp. 952-962, 1999.

[6] Adluru, G., and DiBella, E.V.R. "Reordering for improved constrained recon-struction from undersampled k-space data," Int. J. Biomedical Imaging, vol. 2008, p. 341684, 2008.

[7] Wu, B., Millane, R.P., Watts, R., and Bones P.J. "Applying compressed sensing in parallel MRI," Proc. 16th Ann. Meet. ISMRM, Toronto, p. 1480, 2008.

[8] Wu, B., Bones, P.J., Millane, R.P., and Watts, R. "Prior estimated based compressed sensing in contrast enhanced MRA," Proc. 18th Ann. Meet. ISMRM, Stockholm, 2010.

[9] Wu, B. "Exploiting Data Sparsity in Parallel Magnetic Resonance Imaging," PhD thesis, University of Canterbury, Christchurch, New Zealand, 2009.

[10] Davenport, R. "Derivation of the gamma-variate relationship for tracer dilution curves," J. Nucl. Med., vol. 24, pp. 945–948, 1983.

[11] Madsen, M.T. "A simplified formulation of the gamma variate function," Phys. Med. Biol., vol. 37, pp. 1597-1600, 1992.

[12] Jain, A.K., "Fundamentals of digital image processing," Prentice-Hall, 1989.

[13] Aggarwal, N., Bandyopadhyay, S. and Bresler, Y. "Spatio-temporal modeling and adaptive acquisition for cardiac MRI," Proc. IEEE Int. Conf. Biomed. Eng., pp. 628-631, 2004.

[14] Madore, B., Glover, G.H. and Pelc, N.J. "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD) applied to cardiac imaging and fMRI," Mag. Res. Med., vol. 42, pp 813-828, 1999.