

Bayesian Parallel Imaging with Edge-Preserving Priors

Ashish Raj¹, Gurmeet Singh², Ramin Zabih^{3,4}, Bryan Kressler⁴, Yi Wang⁴, Norbert Schuff¹,
Michael Weiner¹

¹Department of Radiology, University of California at San Francisco and Veterans Affairs
Medical Center, San Francisco

²Department of Electrical Engineering, Cornell University, Ithaca, NY

³Department of Computer Science, Cornell University, Ithaca, NY

⁴Department of Radiology, Weill Medical College of Cornell University, New York, NY

* Corresponding author:

Ashish Raj

Center for Imaging of Neurodegenerative Diseases (CIND)

VA Medical Center (114M)

4150 Clement St

San Francisco, CA 94121

ashish.raj@ucsf.edu

phone : (415) 221 4810 ext 4800

Word Count: 9000

Submitted: November 2, 2005

Revised: June 1, 2006

Bayesian Parallel Imaging with Edge-Preserving Priors

ABSTRACT

Existing parallel magnetic resonance imaging methods are limited by a fundamental trade-off, where suppressing noise introduces aliasing artifacts. Bayesian methods with an appropriately chosen image prior offer a promising alternative; however, previous methods with spatial priors assume that intensities vary smoothly over the entire image, and therefore blur edges. An edge-preserving prior is introduced which instead assumes that intensities are piecewise smooth, and a new approach is proposed to efficiently compute its Bayesian estimate. The estimation task is formulated as an optimization problem, which requires minimizing a non-convex objective function in a space with thousands of dimensions. As a result, traditional continuous minimization methods cannot be applied. This optimization task is closely related to some problems in computer vision for which discrete optimization methods have been developed in the last few years. We adapt these algorithms, which are based on graph cuts, to address our optimization problem. Results of several parallel imaging experiments of brain and torso regions performed under challenging conditions with high acceleration factors are shown and compared with conventional SENSE methods. An empirical analysis indicates visually improved overall quality compared to conventional methods.

Keywords: Parallel imaging, edge-preserving priors, Bayesian reconstruction, SENSE, graph cuts, regularization.

INTRODUCTION

THE use of multiple coils in MR imaging to reduce scan time (and thus motion artifacts) has become quite popular recently. These parallel imaging techniques include, *inter alia*, SENSE (1), (2), (3), SMASH (4), (5) and GRAPPA (6). Good comparative reviews were presented in (7). While all of these methods are mathematically similar, SENSE is the reconstruction method that performs exact matrix inversion (8) and will be the focus of this work. These schemes use multiple coils to reconstruct (unfold) the un-aliased image from under-sampled data in Fourier- or k-space. Successful unfolding relies on receiver diversity - each coil “sees” a slightly different image, since their spatial sensitivity profiles are different.

Unfortunately, the conditioning of the encoding system becomes progressively worse with increasing acceleration factors. Therefore conventional parallel imaging methods, especially at accelerations above 3, suffer from a fundamental noise limitation whereby unfolding is achieved at the cost of noise amplification. This effect depends on coil geometry and the acceleration factor, and is best captured in terms of the g-map, which is a spatial mapping of the noise amplification factor. Reconstructed data can be further degraded in practice by inconsistencies between encoding and decoding sensitivity due to physiological motion, misalignment of coils, and insufficient resolution of sensitivity calibration lines. In this paper we propose a novel Bayesian approach whereby an edge-preserving spatial prior is introduced to reduce noise and improve the unfolding performance of parallel imaging reconstruction. The resulting estimation task is formulated as an optimization problem, whose solution is efficiently obtained by graph-based algorithms.

Proposals to reduce noise and artifacts can be grouped into two classes: attempts to handle sensitivity errors, using a maximum-likelihood approach (9) or Total Least Squares (10); and techniques that exploit some prior information about the imaging target via regularization. While regularization is generally effective for solving ill-posed problems, existing methods rarely exploit the spatial dependencies between pixels. Most techniques either impose minimum norm solutions (as in regularized SENSE (11), (12), (13)), or require a prior estimate of the target. For multi-frame imaging, temporal priors were reported in (14), (15), (16), and the generalized series model was reported using RIGR (17), (18). The limitations of these regularization techniques are clear: regularized SENSE makes unrealistic assumptions about the image norm, while methods relying on a prior estimate of the imaging target (called the mean or reference image) -- like (12), (13), (17), (18) -- must be carefully registered to the target. In practice, the use of such strong reference priors is vulnerable to errors in their estimation, leading to reconstruction

artifacts. Temporal priors are obviously restricted to dynamic imaging. Even though the minimum norm prior reduces noise, it is unsatisfactory for de-aliasing in parallel imaging since it can be shown by simple algebra that it favours solution with equally strong aliases. For example, suppose we have an under-determined aliasing system $y = x_1 + x_2$. Then the minimum norm solution is $x_1 = x_2 = y/2$, which amounts to alias energy being equal to the desired signal energy. This is why conventional regularization techniques are good at reducing noise but counter-effective for removing aliasing. The introduction of spatial priors is essential for this latter task. This is possible within the Tikhonov regularization framework (13), but assumes that intensities vary smoothly over the entire imaging target.

Our approach uses a spatial prior on the image that makes much more realistic assumptions regarding smoothness. Our prior model is quite general, and has very few parameters; hence little or no effort is required to find this prior, in contrast to image-based or temporal priors. The primary challenge in our formulation is a computational one: unlike regularized SENSE there is no closed-form solution, and we need to minimize a non-convex objective function in a space with thousands of dimensions. However, we have developed an efficient algorithm to solve this problem, by relying on some powerful discrete optimization techniques that have been recently developed in the computer vision community (19), (20), (39).

We apply an Edge-Preserving Prior (EPP), which assumes that voxel intensity varies slowly within regions, but – in contrast to smoothness-enforcing Tikhonov regularization – can change discontinuously across object boundaries (21). Since our piecewise smooth model only imposes relationships between neighbouring voxels, it can be used for a very wide range of images (for instance, it is applicable to any MR image, regardless of contrast or modality). EPP's have been studied quite extensively in statistics (22), computer vision (21), (23), (24) and image processing (25), and are widely considered natural image models.

The computational challenge of EPP's is too difficult for conventional minimization algorithms like conjugate gradients or steepest descent. However, EPP's have become widely used in computer vision in the last few years, due primarily to the development of powerful optimization techniques based on graph cuts (19), (20). Graph cuts are discrete methods, where the optimization task is reformulated as the problem of finding a minimum cut on an appropriately constructed graph¹. The minimum cut problem, in turn, can be solved very efficiently by modern graph algorithms (26).

¹ A graph is a set of objects called nodes joined by links called edges. The relationship between graph cuts and optimization problems is described in more detail at the end of this section, and in (19), (20).

While graph cut algorithms can only be applied to a restricted set of problems (19), they have proven to be extremely effective for applications such as stereo matching (20), where they form the basis for most of the top-performing algorithms (27). Although standard graph cut algorithms (19, 20) cannot be directly applied to minimize our objective function, we have developed a graph cut reconstruction technique based on a subroutine due to Hammer (36) that has proven to be quite promising. Preliminary parallel imaging results indicate the potential and promise of this algorithm, which we call EPIGRAM (Edge-preserving Parallel Imaging with GRAPh cut Minimization).

THEORY

Summary of Parallel Imaging

Suppose the target image is given by $X(\mathbf{r})$, where \mathbf{r} is the 2-D spatial vector, \mathbf{k} is a point in k-space. The imaging process for each coil l (from L coils) is given by

$$\mathcal{Y}_l(\mathbf{k}) = \int d\mathbf{r} e^{-i2\pi\mathbf{r}\mathbf{k}} S_l(\mathbf{r}) X(\mathbf{r}) \quad [1]$$

where \mathcal{Y}_l is the (under-sampled) k-space data seen by the l -th coil, and S_l is its sensitivity response. For Cartesian sampling it is well known that [1] reduces to folding in image space, whereby each pixel p in Y_l , the Fourier Transform of \mathcal{Y}_l , results from a weighted sum of aliasing pixels in X . If phase encoding steps are reduced by R times, then the $N \times M$ image folds over into $N/R \times M$ aliased coil outputs. This can be easily verified by discretizing [1] and taking the Fourier transform. For $i = 1, \dots, N$; $\bar{i} = 1, \dots, N/R$; $j = 1, \dots, M$; define $p = (i, j)$, $\bar{p} = (\bar{i}, j)$. Then we have

$$Y_l(\bar{p}) = \sum_{[p] = \bar{p}} S_l(p) X(p) \quad [2]$$

where we denote $[p] = (\text{mod}(i, N/R), j)$. Note that as defined, \bar{p} spans the pixels of Y_l and p the pixels of X . This process is depicted in Figure 1. Over the entire image this has a linear form

$$\mathbf{y} = \mathbf{E} \mathbf{x} \quad [3]$$

where vector $\mathbf{x} = \{x_p \mid p \in \mathcal{P}\}$ is a discrete representation of the intensities of the target image $X(\mathbf{r})$, p indexes the set \mathcal{P} of all pixels of the image, and vector \mathbf{y} contains the aliased images “seen” by the

receiver coils. Matrix \mathbf{E} encodes coil sensitivity responses, and is a $L \times R$ block-diagonal matrix of the form shown in Figure 2.

SENSE takes a least squares approach via the pseudo-inverse of \mathbf{E} :

$$\hat{\mathbf{x}}_{SENSE} = (\mathbf{E}^H \mathbf{E})^{-1} \mathbf{E}^H \mathbf{y} \quad [4]$$

This is the maximum likelihood estimate under the assumption of additive white Gaussian noise (28, Ch. 15). Unfortunately inverse problems of this form become progressively ill-posed with increasing acceleration, leading to noise amplification and insufficient de-aliasing in many cases. In order to reduce these effects and stabilize the inversion in SENSE, Tikhonov-type regularization was introduced (10), (13):

$$\hat{\mathbf{x}}_{regSENSE} = \arg \min_{\mathbf{x}} \left\{ \|\mathbf{E}\mathbf{x} - \mathbf{y}\|^2 + \mu^2 \|\mathbf{A}(\mathbf{x} - \mathbf{x}_r)\|^2 \right\} \quad [5]$$

where the first term enforces agreement with observed data, and the second penalizes non-smooth solutions through an appropriate matrix \mathbf{A} and some prior reference image \mathbf{x}_r . Its closed-form solution is

$$\hat{\mathbf{x}}_{regSENSE} = \mathbf{x}_r + (\mathbf{E}^H \mathbf{E} + \mu^2 \mathbf{A}^H \mathbf{A})^{-1} \mathbf{E}^H (\mathbf{y} - \mathbf{E}\mathbf{x}_r) \quad [6]$$

Both [4] and [6] can be computed very quickly for the Cartesian case since they readily break up into independent $L \times R$ sub-problems (for the standard choice of $\mathbf{A} = \mathbf{I}$, the identity matrix). If there is no reference image (i.e., $\mathbf{x}_r = 0$), then with $\mathbf{A} = \mathbf{I}$ this computes the minimum norm solution (11), while more general Tikhonov forms of \mathbf{A} impose global smoothness.

Bayesian Reconstruction

Even with regularization there is a noise/unfolding limit. If μ is too small, there will be insufficient noise reduction. If μ is too high, noise will be removed but residual aliasing will occur. This fundamental aliasing/noise limit cannot be overcome unless more information about the data is exploited. This naturally suggests a Bayesian approach, a subject of some recent work (29). Given the imaging process [3], observation \mathbf{y} and the prior probability distribution $Pr(\mathbf{x})$ of the target image \mathbf{x} , Bayesian methods maximize the posterior probability

$$Pr(\mathbf{x}/\mathbf{y}) \propto Pr(\mathbf{y}/\mathbf{x}) \cdot Pr(\mathbf{x}) \quad [7]$$

The first right-hand term, called the likelihood function, comes from the imaging model [3], and the second term is the prior distribution. In the absence of a prior, this reduces to maximizing the likelihood, as performed by SENSE. Assuming that $\mathbf{n} = \mathbf{y} - \mathbf{E}\mathbf{x}$ is white Gaussian noise, [3] implies a simple Gaussian distribution for $Pr(\mathbf{y}/\mathbf{x})$:

$$Pr(\mathbf{y}/\mathbf{x}) \propto e^{-\|\mathbf{y}-\mathbf{E}\mathbf{x}\|^2} \quad [8]$$

Similarly, we can write the prior, without loss of generality, as

$$Pr(\mathbf{x}) \propto e^{-G(\mathbf{x})} \quad [9]$$

Depending on the term $G(\mathbf{x})$, this form can succinctly express both the traditional Gaussianity and/or smoothness assumptions, as well as more complicated but powerful Gaussian or Gibbsian priors modeled by Markov Random Fields (MRFs) (23, 24). The posterior is maximized by

$$\hat{\mathbf{x}}_{MAP} = \arg \min_{\mathbf{x}} \left(\|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2 + G(\mathbf{x}) \right), \quad [10]$$

which is the maximum *a posteriori* (MAP) estimate. Conventional image priors impose spatial smoothness; hence this can be viewed as the sum of a data penalty and a smoothness penalty. The data penalty forces \mathbf{x} to be compatible with the observed data, and the smoothness penalty $G(\mathbf{x})$ penalizes solutions that lack smoothness. Traditionally, only smoothness penalties of the kind $G(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2$ have been used, where \mathbf{A} is a linear differential operator. This corresponds to Eqn. [6] if the reference image $\mathbf{x}_r = 0$, and is commonly known as Tikhonov regularization (28).

However, this smoothness penalty assumes that intensities vary smoothly across the entire image. Such an assumption is inappropriate for most images, since most image data change smoothly, but have discontinuities at object boundaries. As a result, Tikhonov smoothness penalty causes excessive edge blurring, while we seek an edge-preserving G . To illustrate this, we show in Figure 3 a single noisy image row, and two possible strategies to de-noise it. The difference in performance between global smoothing (obtained by Gaussian blurring) and edge-preserving smoothing (via median filtering) is obvious: whereas both denoise the signal, one over-smoothes sharp transitions, but the other largely preserves them.

Edge-Preserving Priors in MR Imaging

A natural class of edge-preserving smoothness penalties is

$$G_{EP}(\mathbf{x}) = \sum_{(p,q) \in \mathcal{N}_s} V(x_p, x_q) \quad [11]$$

The spatial neighbourhood system \mathcal{N}_s consists of pairs of adjacent pixels, usually the 8-connected neighbours. The *separation cost* $V(x_p, x_q)$ gives the cost to assign intensities x_p and x_q to neighbouring pixels p and q ; the form of this prior can be justified in terms of Markov Random Fields (23). Typically V has a non-convex form such as $V(x_p, x_q) = \lambda \min(|x_p - x_q|, K)$ for some metric $|\cdot|$ and constants K, λ . Such functions effectively assume that the image is piecewise smooth rather than globally smooth. Figure 4 shows two possible choices of V ; the left one does not preserve edges, while the right one does. For MR data the truncated linear model appears to work best; it seems to present the best balance between noise suppression (due to the linear part) as well as edge-preservation (due to truncation of penalty function). Therefore, neighbouring intensity differences within the threshold K will be treated as noise and penalized accordingly. However, larger differences will not be further penalized, since they occur, most likely, from the voxels being separated by an edge. Note that this is very different from using a traditional convex distance for the separation cost, such as the L_2 norm, which effectively forbids two adjacent pixels from having very different intensities. Although the L_2 separation cost does not preserve edges, it is widely used since its convex nature vastly simplifies the optimization problem.

A possible problem with the truncated linear penalty is that it can lead to some loss of texture, since the Bayesian estimate will favour images with piecewise smooth areas rather than textured areas. In Discussion we point out examples of this feature and suggest ways to mitigate it.

Parallel Imaging As Optimization

The computational problem we face is to efficiently minimize

$$\|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2 + G_{EP}(\mathbf{x}) \quad [12]$$

From (4) we know that \mathbf{E} has a diagonal block structure, and decomposes into separate interactions between R aliasing voxels, according to Eq. [2]. Let us first define for each pixel $\bar{p} = (\bar{i}, j)$ in Y_l the set

of aliasing pixels in X that contribute to $Y_l(\bar{p})$, as follows: For image X of size $M \times N$ undergoing R -fold acceleration, aliasing only occurs in the phase encode direction, between aliasing pixels. Then

$$\|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2 = \sum_{\bar{p}} \sum_l \left[Y_l(\bar{p}) - \sum_{[p]=\bar{p}} S_l(p) x_p \right]^2 \quad [13]$$

This can be intuitively understood by examining the aliasing process depicted in Figure 1. After some rearrangement, this expands to

$$\begin{aligned} \|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2 = & \sum_{\bar{p}} \sum_l Y_l^2(\bar{p}) + \sum_p \left(\sum_l S_l^2(p) \right) x_p^2 \\ & - 2 \sum_p \left(\sum_l S_l(p) Y_l([p]) \right) x_p + 2 \sum_{(p,p') \in \mathcal{N}_a} \left(\sum_l S_l(p) S_l(p') \right) x_p x_{p'}, \end{aligned} \quad [14]$$

where we define the aliasing neighbourhood set $\mathcal{N}_a = \{(p, p'), [p] = [p'], p \neq p'\}$ over all aliasing pairs. Grouping terms under single pixel and pairwise interactions, we get

$$\|\mathbf{y} - \mathbf{E}\mathbf{x}\|^2 = a^2 + \sum_p b(p) x_p^2 - 2 \sum_p c(p) x_p + 2 \sum_{(p,p') \in \mathcal{N}_a} d(p, p') x_p x_{p'},$$

for appropriately chosen functions $b(p)$, $c(p)$ and $d(p, p')$. The first term is a constant and can be removed from the objective function; the next two terms depend only on a single pixel while the last term depends on two pixels at once, both from the aliasing set. This last term, which we will refer to as a *cross term*, arises due to the non-diagonal form of our system matrix \mathbf{E} .

To perform edge-preserving parallel imaging, we need to minimize our objective function:

$$\mathcal{E}(\mathbf{x}) = a^2 + \sum_p b(p) x_p^2 - 2 \sum_p c(p) x_p + 2 \sum_{(p,p') \in \mathcal{N}_a} d(p, p') x_p x_{p'} + \sum_{(p,q) \in \mathcal{N}_s} V(x_p, x_q) \quad [15]$$

Let us first consider the simpler case of our objective function that would arise if \mathbf{E} were diagonal. In this case there would be no cross terms (i.e., $d(p, p') = 0$), which appears to simplify the problem considerably. Yet even this simplification results in a difficult optimization problem. There is no closed form solution, the objective function is highly non-convex, and the space we are minimizing over has thousands of

dimensions (one dimension per pixel). Worse still, minimizing such an objective function is almost certain to require an exponentially large number of steps².

If \mathbf{E} were diagonal, however, the objective function would be in a form that has been extensively studied in computer vision (10) (20) (27), and where significant recent progress has been made. Specifically, a number of powerful methods have been designed for such problems, using a discrete optimization technique called graph cuts (20), which we briefly summarize in the next section. Graph cuts are a powerful method to minimize $\mathcal{E}(\mathbf{x})$ in [15], and can be easily applied as long as \mathbf{E} is diagonal (19). The presence of off-diagonal entries in \mathbf{E} gives rise to cross terms in our objective function, making traditional graph cut algorithms (19,20) inapplicable, and requiring an extension described in the Method section.

Optimization with Graph Cuts

Objective functions similar to [15] can be minimized by computing the minimum cut in an appropriately defined graph (“graph cuts”). This technique was first used for images in (30), who used it to optimally denoise binary images. A recent series of papers (19) (20) extended the method significantly, so that it can now be used for problems such as stereo matching (20), (31) and image/video synthesis (32) in computer vision; medical image segmentation (33); and fMRI data analysis (34).

The basic idea is to first discretize the continuous pixel intensities x_p into a finite discrete set of *labels* $\mathcal{L} = \{1, \dots, N_{\text{labels}}\}$. Since we focus on MR reconstruction problems, we will assume labels are always intensities, and use the terms interchangeably; however, graph cut algorithms are used on a wide variety of problems in computer vision and graphics, and often use labels with a more complex meaning. Then instead of minimizing over continuous variables x_p we minimize over individual labels $\alpha \in \mathcal{L}$, allowing any pixel in the image to take the label α . In practice the dynamic range of intensities may need to be reduced for computational purposes, although this is not a requirement of our technique. The most powerful graph cut method is based upon expansion moves. Given a labeling $\mathbf{x} = \{x_p \mid p \in \mathcal{P}\}$ and a label α , an α -expansion $\chi = \{\chi_p \mid p \in \mathcal{P}\}$ is a new labeling where χ_p is either x_p or α . Intuitively, χ is constructed from \mathbf{x} by giving some set of pixels the label α . The expansion move algorithm picks a label α , finds the lowest cost χ and moves there. This is pictorially depicted in Figure 5.

² More precisely, it was shown in (20) to be NP-hard, which means that it is in a class of problems that are widely believed to require exponential time in the worst case.

The algorithm converges to a labeling where there is no α -expansion that reduces the value of the objective function \mathcal{E} for any α . The key subroutine in the expansion move algorithm is to compute the α -expansion χ that minimizes \mathcal{E} . This can be viewed as an optimization problem over binary variables, since during an α -expansion each pixel either keeps its old label or moves to the new label α . This is also shown in Figure 5. An α -expansion χ is equivalent to a binary labeling $\mathbf{b} = \{b_p \mid p \in \mathcal{P}\}$ where

$$\chi_p = \begin{cases} x_p & \text{iff } b_p = 0 \\ \alpha & \text{iff } b_p = 1 \end{cases} \quad [16]$$

Just as for a labeling χ there is an objective function \mathcal{E} , for a binary labeling \mathbf{b} there is an objective function B . More precisely, assuming χ is equivalent to \mathbf{b} , we define B by

$$B(\mathbf{b}) = \mathcal{E}(\chi)$$

We have dropped the arguments \mathbf{x} , α for clarity, but the equivalence between the α -expansion χ and the binary labeling \mathbf{b} clearly depends on the initial labeling \mathbf{x} and on α .

In summary, the problem of computing the α -expansion that minimizes \mathcal{E} is equivalent to finding the \mathbf{b} that minimizes the binary objective function B . The exact form of B will depend on \mathcal{E} . Minimization of \mathcal{E} proceeds via successive binary minimizations corresponding to expansion moves. The binary minimization subroutine is somewhat analogous to the role of line search in Conjugate Gradient, where a local minimum is repeatedly computed over different 1D search spaces. With graph cuts however, the binary subroutine efficiently computes the *global* minimum over $2^{|\mathcal{P}|}$ candidate solutions, where $|\mathcal{P}|$ is the number of pixels in the image. In contrast to traditional minimization algorithms like Conjugate Gradients, trust region, simulated annealing etc (28), graph cuts can therefore efficiently optimize highly non-convex objective functions arising from edge-preserving penalties (19).

Consider a binary objective function of the form

$$B(\mathbf{b}) = \sum_p B_1(b_p) + \sum_{p,q} B_2(b_p, b_q) \quad [17]$$

Here, B_1 and B_2 are functions of binary variables; the difference is that B_1 depends on a single pixel, while B_2 depends on pairs of pixels. Graph cut methods minimize B by reducing computing a minimum cut on an appropriately constructed graph. The graph consists of nodes which are voxels of the image as well as

two special terminal nodes, as shown in Figure 6. The voxel nodes have been labeled p, q, r etc and terminal nodes by S, T . All nodes are connected to both terminals via edges, each of which have weights obtained from the B_1 terms above. Nodes are also connected to each other via edges with weights obtained from the pairwise interaction term B_2 .

The binary optimization problem can be solved by finding the minimum cut on this graph (20). A cut is defined as a partition of the graph into two connected subgraphs, each of which contains one terminal. The minimum cut minimizes the sum of the weights of the edges between the subgraphs. There are fast algorithms to find the minimum cut, using max-flow methods (26). It is shown in (19) that the class of B that can be minimized exactly by computing a minimum cut on such a graph satisfies the condition

$$B_2(0,0)+B_2(1,1) \leq B_2(1,0)+B_2(0,1) \quad [18]$$

If $B_2(x, y)$ satisfies [18], then it is said to be *submodular* with respect to x and y , and a function B is called submodular if it consists entirely of submodular terms.³ Single-variable terms of the form of B_1 are always submodular. We will refer to the set of all pixel pairs for which $B_2(b_p, b_q)$ are submodular as the submodular set \mathcal{S} .

Previous applications of graph cuts have been for diagonal \mathbf{E} ; this leads to no cross terms, and so B_1 comes solely from the data penalty and B_2 comes only from the smoothness penalty. It is shown in (20) that if the separation cost is a metric then B_2 satisfies Eq. [18]. Many edge-preserving separation costs are metrics, including the truncated cost V used here and shown in Figure 2(b) (see (20) for details). However, the situation is more complicated in parallel MR reconstruction, where \mathbf{E} is non-diagonal. As a result the data penalty has pairwise interactions due to the presence of the cross-terms $d(p, p')$ in Eq. [15]. This also follows from Figure 1, which shows how this data penalty arises from the joint effect of both aliasing voxels p and p' in the image.

It was previously shown (35) that the binary optimization problem arising from [15] is in general submodular only for a small subset of all cross terms. This necessitates the use of a subroutine due to (36) to accommodate cross-terms arising in MR reconstruction, as described in the next section.

³ Some authors (21) use the term regular instead of submodular.

METHODS

A New MR Reconstruction Algorithm Based On Graph Cuts

The subroutine we use to find a good expansion move is closely related to relaxation methods for solving integer programming problems (26), where if the linear programming solution obeys the integer constraints it solves the original integer problem. We compute an expansion move by applying the algorithm of Hammer *et. al.* (36), which was introduced into computer vision in early 2005 by Kolmogorov (39). The theoretical analysis and error bounds presented in (40) help explain the strong performance of this construction for MR reconstruction.

For each pixel p , we have a binary variable b_p that was 1 if p acquired the new label α and 0 otherwise. We will introduce a new binary variable \tilde{b}_p , which will have the opposite interpretation (i.e., it will be 0 if p acquired the new label α and 1 otherwise). We will call a pixel *consistent* if $\tilde{b}_p = 1 - b_p$. Instead of our original objective function $B(\mathbf{b})$ we minimize a new objective function $\tilde{B}(\mathbf{b}, \tilde{\mathbf{b}})$, where $\tilde{\mathbf{b}}$ is the set of new binary variables $\tilde{\mathbf{b}} = \{\tilde{b}_p \mid p \in \mathcal{P}\}$. $\tilde{B}(\mathbf{b}, \tilde{\mathbf{b}})$ is constructed so that $\tilde{\mathbf{b}} = \mathbf{1} - \mathbf{b} \Rightarrow \tilde{B}(\mathbf{b}, \tilde{\mathbf{b}}) = B(\mathbf{b})$ (in other words, if every pixel is consistent then the new objective function is the same as the old one). Specifically, we define our new objective function by

$$\begin{aligned} 2 \cdot \tilde{B}(\mathbf{b}, \tilde{\mathbf{b}}) = & \sum_p \left(B_1(b_p) + B_1(1 - \tilde{b}_p) \right) + \sum_{(p,q) \in \mathcal{S}} \left(B_2(b_p, b_q) + B_2(1 - \tilde{b}_p, 1 - \tilde{b}_q) \right) \\ & + \sum_{(p,q) \notin \mathcal{S}} \left(B_2(b_p, 1 - \tilde{b}_q) + B_2(1 - \tilde{b}_p, b_q) \right) \end{aligned} \quad [19]$$

Here, the functions $B_1(\cdot)$ and $B_2(\cdot)$ come from our original objective function B in [17].

Importantly, our new objective function $\tilde{B}(\mathbf{b}, \tilde{\mathbf{b}})$ is submodular. The first summation only involves $B_1(\cdot)$, while for the remaining two terms, simple algebra shows that

$$\begin{aligned} B_2(b, b') \text{ is submodular} & \Rightarrow B_2(1 - b, 1 - b') \text{ is submodular} \\ B_2(b, b') \text{ is non-submodular} & \Rightarrow \text{both } B_2(b, 1 - b') \text{ and } B_2(1 - b, b') \text{ are submodular} \end{aligned}$$

As a result, the last two summations in [19] only contain submodular terms. Thus $\tilde{B}(\mathbf{b}, \tilde{\mathbf{b}})$ is submodular, and can be easily minimized using the binary graph cut subroutine. In summary, minimizing $\tilde{B}(\mathbf{b}, \tilde{\mathbf{b}})$ is exactly equivalent to minimizing our original objective function B , as long as we obtain a solution where every pixel is consistent. We note that our technique is not specific to MR reconstruction, but can compute the MAP estimate of an arbitrary linear inverse system under an edge-preserving prior G_{EP} .

While we cannot guarantee that all pixels are consistent, in practice this is true for the vast majority of pixels (typically well over 95%). In our algorithm, we simply allow pixels that are not consistent to keep their original labels rather than acquiring the new label α . However, even if there are pixels which are not consistent, this subroutine has some interesting optimality properties. It is shown in (36) that any pixel which is consistent is assigned its optimal label. As a result, our algorithm finds the optimal expansion move for the vast majority of pixels.

Convergence properties

We investigated the convergence properties of the proposed technique using simulated data from a Shepp-Logan phantom, with intensities quantized to integer values between 0 and 255. We computed the objective function [15] achieved after each iteration, for 3x acceleration and 8 coils.

In vivo experiments : Setup and Parameters

High field strength (4 Tesla) structural MRI brain data was obtained using a whole body scanner (Bruker/Siemens Germany) equipped with a standard birdcage, 8 channel phased-array transmit/receive head-coil localized cylindrically around the S-I axis. Volumetric T1-weighted images (1 x 1 x 1 mm³ resolution) were acquired using a Magnetization-Prepared RAPid Gradient Echo (MPRAGE) sequence with TI/TR = 950/2300 ms timing and a flip angle of 8°. Total acquisition for an unaccelerated dataset is about 8:00 minutes. In a separate study, images of the torso region were acquired using a Gradient Echo sequence with a flip angle of 60 degrees and TE/TR of 3.3/7.5 ms on a GE 1.5T Excite-11 system. Several axial and oblique slices of full resolution data (256 x 256) were acquired from an 8 channel upper body coil arranged cylindrically around the torso.

In order to allow quantitative and qualitative performance evaluations, we acquired all data at full resolution and no acceleration. The aliased images for acceleration factors between 3 and 5 were

obtained by manually undersampling in k-space. In each case the full rooted sum of squares (RSOS) image was also computed after dividing by the relative sensitivity maps obtained from calibration lines. We used the self-calibrating strategy for sensitivity estimation, whereby the centre of k-space is acquired at full density and used to estimate low-frequency (relative) sensitivity maps. We used the central 40 densely sampled calibration lines for this purpose. These lines were multiplied by an appropriate Kaiser-Bessel window to reduce ringing and noise, zero-padded to full resolution and transformed to the image domain. Relative sensitivity was estimated by dividing these images by their RSOS. To avoid division by zero a small threshold was introduced in the denominator, amounting to 5 percent of maximum intensity. This also served effectively to make the sensitivity maps have zero signal in background regions. However, further attempts at segmenting background/foreground from this low-frequency data proved unreliable in some cases and we did not implement it.

Algorithmic parameters were chosen empirically. It was sufficient to quantize intensity labels to $N_{labels} = 256$, since the resulting quantization error is much smaller than observed noise. The computational cost of EPIGRAM grows linearly with N_{labels} , so fewer labels are preferable. Model parameters were varied (geometrically) over the range $K \in [N_{labels} / 20, N_{labels} / 2]$, $\lambda \in [0.01 \cdot \max(x), 1 \cdot \max(x)]$ to find the best values. However, performance was found to be rather insensitive to these choices, and we used the same parameters for all cases shown in this paper. Graph cut algorithms are typically insensitive to initialization issues, and we chose the zero image as an initial guess. All reconstructions were obtained after twenty iterations. Regularized SENSE reconstruction was used to compare with our method. Regularization factor μ was chosen after visually evaluating image quality with a large range of values in the region $\mu \in [0.01, 0.6]$. The images giving the best results are shown in the next section, along with those obtained with a higher regularization. The latter result was obtained to observe the noise versus aliasing performance of SENSE.

Quantitative Performance Evaluation

Apart from visual evidence presented in the next section, a quantitative performance evaluation was conducted on reconstructed *in vivo* data. For *in vivo* data the problem of ascertaining noise estimate or other performance measures is challenging due to the non-availability of an accurate reference image. Unfortunately none of the reconstruction methods we implemented, whether RSOS, regularized SENSE or EPIGRAM, are unbiased estimators of the target. This makes direct estimation of noise performance difficult, and the traditional root mean square error (RMSE) is inadequate. We follow instead a recent

evaluation measure for parallel imaging methods proposed by Reeder et al (37), which provides an unambiguous and fair comparison of SNR and geometry factor. Two separate scans of the same target with identical settings are acquired, and their sum and difference obtained. The local signal level at each voxel is computed by averaging the sum image over a local window, and the noise level obtained from the standard deviation of the difference image over the same window. Then

$$SNR = \frac{\text{mean}(\text{Sum image})}{\sqrt{2} \text{stdev}(\text{Diff image})}$$

Here the mean and standard deviation are understood to be over a local window, in our case a 5x5 window around the voxel in question. This provides unbiased estimates that are directly comparable across different reconstruction methods. We perform a similar calculation, with a crucial difference: instead of acquiring two scans, we use a single scan but add random uncorrelated Gaussian noise in the coil outputs to obtain two noisy data sets. This halves the acquisition effort without compromising estimate quality, since uncorrelated noise is essentially what the two-scan method also measures. Reeder et al explain that their method can be erroneous for *in vivo* data because motion and other physiological effects can seriously degrade noise estimates. Our modification, while achieving the same purpose, does not suffer from this problem, and hence should be more appropriate for *in vivo* data. The SNR calculation also allows the geometry factor or g-factor maps for each reconstruction method to be obtained. For each voxel p and acceleration factor R :

$$g(p, R) = \frac{SNR_{RSOS}(p)}{SNR(p)\sqrt{R}}$$

where SNR_{RSOS} is the SNR of the RSOS reconstruction. The SNR and g-map images appeared quite noisy due to the division step and estimation errors, and we had to smoothen them for better visualization.

Comparing EPIGRAM and Regularized SENSE

SNR and g-factor maps obtained by the method of (37) and above are well-suited for unbiased reconstructions like unregularized SENSE since they provide comparable measures of noise amplification. Reeder et al demonstrate conclusively that the two-scan method gives comparable estimates to those obtained voxel-wise by using hundreds of identical scans. Unfortunately, neither SNR nor g-factor can act as a single measure of performance in the case of biased reconstructions like regularized SENSE, EPIGRAM, or indeed any Bayesian estimate. This is because now there are two

quite different sources of degradation: noise amplification, as well as aliasing. While the former can be measured accurately by SNR and g-maps using the above technique, the latter cannot. All regularization or Bayesian approaches come with external parameters, which can be thought of as “knobs” available to the user - μ for SENSE and λ for EPIGRAM. Depending on how much the user tweaks those knobs, it is possible to achieve any desired degree of noise performance. To illustrate this, we show in Results the effect of μ , the regularization parameter in SENSE. We demonstrate that any desired g-factor can be achieved by making μ large enough. Therefore, for a proper comparison of reconstruction methods, it is important to fix either aliasing quality or noise quality, and then compare the other measure. In this section we adopt the following approach: we turn the “knobs” until both EPIGRAM and regularized SENSE produce approximately similar g-maps and mean g values. Then we visually evaluate how each method performed in terms of aliasing and overall reconstruction quality. In each case we tabulate the mean SNR and g values achieved by the given settings. These are contained in Results.

RESULTS

Convergence result

The convergence behaviour of the objective function [15] against the number of iterations for $R=3$, $L=8$ is shown in Figure 7. As is typical with graph cut algorithms (31), most of the improvement happens in the first few (approx 5) iterations. In our initial implementation of EPIGRAM, which is written primarily in MatLab, an iteration takes about a minute on this data set. Since EPIGRAM is almost linear in the number of nodes, total running time approximately scales as N^2 , the image size.

In vivo Results

The best parameter values were consistent across all *in vivo* data sets we tried: $K = N_{\text{labels}} / 7$, $\lambda = 0.04 \cdot \max(x)$. Results of several brain imaging experiments with these parameter values are displayed in Figures 8-9. Figure 8 shows reconstruction of a MPRAGE scan of a central sagittal slice, with an undersampling factor $R=4$ along A-P direction. The RSOS reference image is shown in (a), regularized SENSE with (empirically obtained optimal) $\mu = 0.08$ in (b), regularized SENSE with $\mu = 0.16$ in (c) and EPIGRAM in (d). Reduction in noise is visually noticeable in the EPIGRAM reconstruction, compared to both SENSE reconstructions. Higher regularization in SENSE caused unacceptable aliasing, as observed in (c). We note that unregularized (i.e. standard) SENSE results were

always worse than regularized SENSE, and have consequently not been shown. Another sagittal scan result is shown in Figure 9, this time from the left side of the patient. Image support is smaller, allowing 5x acceleration. The optimally regularized SENSE output (b), with $\mu = 0.1$ is noisy at this level of acceleration, and $\mu = 0.2$ in (c) introduced significant aliasing, especially along the central brain region. EPIGRAM (d) exhibits some loss of texture, but on the whole, seems to outperform SENSE.

A set of torso images acquired on GE 1.5T scanner using a GRE sequence and acceleration factor of 3 (along A-P direction) is resolved from 40 sensitivity calibration lines. Various slice orientations, both axial and oblique, were used. Shown in Figure 10, this data also shows the practical limitation of SENSE when an inadequate number of calibration lines are used for sensitivity estimation. Reconstruction quality of SENSE is poor as a result of the combination of ill-conditioning of the matrix inverse and calibration error. SENSE exhibits both high noise as well as residual aliasing. In fact what appears at first sight to be uncorrelated noise is, upon finer visual inspection, found to arise from unresolved aliases, as the background in 10(b) clearly indicates. EPIGRAM has been able to resolve aliasing correctly and suppress noise, without blurring sharp edges and texture boundaries. To demonstrate the performance of these methods more clearly, we show in (d)-(f) zoomed in versions of the images in (a)-(c).

Next we demonstrate the trade-off between noise and aliasing performance. Figure 11 shows another torso slice, along with associated g-maps computed as specified in Methods. We investigate the effect of various regularizations of SENSE. The leftmost column, (a) and (e), show the SENSE result and its g-map for $\mu = 0.1$. Clearly, there is inadequate noise suppression, where some regions of the image have a g-factor as high as 6. It is easy to reduce noise amplification by increasing regularization. In the next column results for $\mu = 0.3$ are shown. The g-map has become correspondingly flatter, but the reconstruction indicates that this has been achieved at the cost of introducing some aliasing in the image. The rightmost column shows EPIGRAM results, which has significantly lower g values, as well as lower aliasing artifacts. In the third column we show that by an appropriate choice of μ it is possible to match the EPIGRAM g-map – compare (g) and (h). However, at this choice of $\mu = 0.5$, SENSE yields unacceptable aliasing. Table 1 shows mean SNR and g-factor values of EPIGRAM and various regularizations of SENSE.

We observe that the regularization needed in SENSE to match the EPIGRAM g values (approximately $\mu = 0.5$ in almost all cases) yields unacceptable aliasing. Instead, a more modest μ must be chosen empirically. The standard RMSE criterion does not really help here. In Figure 12 we plot the average over all our torso data, of the root mean square error for various regularization factors. For this data set,

the optimal μ in terms of RMSE was found to be around 0.25, although in practice the best visual quality was observed below this value, around 0.1-0.15. This seems to suggest that the RMSE measure does not capture reconstruction quality very accurately; in particular it seems to under-emphasize the effect of residual aliasing. In all experimental data shown in this section, the optimal regularization for SENSE was obtained empirically for each data set by visual inspection as far as possible. To give an idea of the visual quality corresponding to a certain μ value, Figure 12 also shows a portion of the resulting SENSE reconstruction. Mean SNR and g-factor for all reconstruction examples presented here are summarized in Table 2. Non-regularized SENSE data are not shown since they are always worse than regularized SENSE.

In all these examples there are regions in EPIGRAM data with some loss of texture, as well as regions of low signal which appear to have been “washed out” by a uniform value. Both effects result from the piecewise smooth assumption imposed by the prior. We discuss ways to mitigate this problem in Discussions, but note that it is typical of most applications where MRF priors have been applied.

DISCUSSION AND CONCLUSIONS

The experimental data shown above indicates the promise of improved performance and visual quality by using EPIGRAM. This is a direct result of our use of a powerful edge-preserving prior model, and the efficient implementation of the associated Bayesian problem using Graph Cut algorithms. An exhaustive clinical evaluation is needed to further characterize the method and identify specific clinical applications – this work is ongoing. It is also desirable to develop appropriate quantitative measures for comparing reconstruction performance of various methods. Conventional measures like SNR and g-factors maps do not well-serve this purpose. We have argued that looking solely at the reduction in noise as a measure of regularization performance is inadequate since it does not account for the introduction of concomitant aliasing artifacts (see for instance Fig 11). In fact g-factor maps on their own might mislead one to conclude that large regularizations are always good since they produce small g-factors. Further work is therefore needed to devise more suitable quantitative measures of performance, for instance a statistically sound diagnostic score obtained blindly from a panel of experts.

We currently do not consider sensitivity errors or temporal information, but such information can be easily incorporated within our objective function. In particular, the use of graph cut minimization for temporal or multi-dimensional data looks very attractive. The graph cut based cost minimization approach should be readily generalizable to other parallel imaging methods like GRAPPA. Since GRAPPA also performs essentially a least squares fit via the optimization of some objective function, its

implementation using graph cut based minimization appears possible. Currently the method is implemented only for Cartesian sampling. Extension of EPIGRAM to non-Cartesian trajectories is an interesting but challenging task.

A shortcoming of our method is the possible loss of texture due to the piecewise smooth assumption. In this work, the model parameters K, λ are determined heuristically and assumed constant over the entire image. Inappropriate choice of either of these parameters can adversely affect reconstructed data. If λ is too large, it will force low signal areas to become uniformly zero, or remove informative texture from non-smooth regions. Both these effects can be observed in our example reconstructions. Decreasing λ can remove this to some extent, but the most effective approach might be an adaptive and locally varying parameter choice. Although our experimentation produced the best results for the chosen truncated linear prior model, other models are possible – truncated quadratic, piece-wise linear, etc. The power of truncated linear model appears well-substantiated by research in computer vision, but future work in MR might conceivably yield better edge-preserving models which preserve textures and low signal regions. Further work on different prior models and their parameters will most likely improve the results shown here. However, it is important to note that the EPIGRAM algorithm can continue to be used in its current form, since it is formulated for arbitrary objective functions of the kind shown in [15].

It is also noteworthy that Bayesian reconstruction can sometimes produce images which appear different, or even better than the best available reconstruction – in our case the RSOS. This is an interesting case of a Bayesian estimate fitting the prior better than the best RSOS estimate. This is a well-known phenomenon in Bayesian estimation, and can be explained by the fact that our MRF prior models the *target* image, not the RSOS reconstruction. Another example of this can be seen in the g-factor maps of EPIGRAM, which shows some regions with $g < 1$. This counter-intuitive phenomenon results from the fact that the prior imposes smoothness to such an extent that the Bayesian estimate becomes less noisy than the unaccelerated image in some places. In the extreme case when $\mu, \lambda = \infty$, both SENSE and EPIGRAM will simply reproduce their respective priors and we will have $g=0$ everywhere! Of course such a result will be of no real use since it is independent of the observed data.

Comparison with Other Prior-driven approaches

It is instructive to compare our EPP model with other prior models. Some very interesting and impressive results for dynamic imaging models (RIGR) were presented in (17), (18) using a powerful generalized

series to model k-space truncation, and dynamic and static parts of MR data. Our Bayesian approach is an interesting counterpoint to this body of work. RIGR imposes a priori constraints in the form of a detailed generalized series model. This approach works best for dynamic imaging, and requires a good, well-registered, high-resolution reference image. This certainly has significant value for many imaging situations, but suffers from high sensitivity to modeling errors, patient motion, as well as the difficulty of obtaining correct model parameters from partial data. Our prior model is simpler and does not need a reference image. The advantage of our approach is that it is basically parameter-free apart from the particular choice of separation cost. Since the EPP model relies solely on spatial coherence, it is effective under widely varying imaging conditions and should be robust against artifacts caused by geometric distortion, B1 intensity variations, excessive noise, and even small amounts of motion. Having said that, it appears possible in theory to incorporate RIGR-type priors within the minimization framework proposed here. Another interesting approach that has the potential to improve conventional regularization schemes is the total variation (TV) method (38). Both EPIGRAM and TV are novel and basically unproven techniques for MR, and it will be interesting to analytically and empirically compare the two.

In conclusion, we have argued that Bayesian estimation with edge-preserving image priors is an effective means of breaking out of the fundamental noise versus aliasing trade-off in conventional parallel imaging without causing over-smoothing. We showed that this Bayesian estimate is a minimization problem that is closely related to problems that have been addressed in the computer vision community. We demonstrated that standard graph cut optimization methods (19),(20) cannot be applied directly to our problem, and instead proposed a new MR reconstruction technique relying on a subroutine due to (36). The resulting MR reconstruction algorithm, which we call EPIGRAM, appears promising both visually and quantitatively. The resulting gains can potentially be exploited to further accelerate acquisition speed in parallel imaging, or to achieve higher reconstruction quality at the same speed.

ACKNOWLEDGMENTS

We thank Vladimir Kolmogorov for spending a considerable amount of time explaining the algorithm of (36); his help was invaluable in clarifying this important work. We also thank Lara Stables and Xiaoping Zhu for help with data acquisition.

REFERENCES

- (1) Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity Encoding For Fast MRI. *Magnetic Resonance in Medicine* 1999; 42(5): 952-962.

- (2) Pruessmann KP, Weiger M, Boernert P, Boesiger P. Advances In Sensitivity Encoding With Arbitrary K-Space Trajectories. *Magnetic Resonance in Medicine* 2001; 46(4):638--651.
- (3) Weiger M, Pruessmann KP, Boesiger P. 2D SENSE For Faster 3D MRI. *Magnetic Resonance Materials in Biology, Physics and Medicine* 2002; 14(1):10-19.
- (4) Sodickson DK, Manning WJ. Simultaneous Acquisition Of Spatial Harmonics (SMASH): Fast Imaging With Radiofrequency Coil Arrays. *Magnetic Resonance in Medicine* 1997; 38(4): 591-603.
- (5) Sodickson DK, McKenzie CA, Ohliger MA, Yeh EN, Price MD. Recent Advances In Image Reconstruction, Coil Sensitivity Calibration, And Coil Array Design For SMASH And Generalized Parallel MRI. *Magnetic Resonance Materials in Biology, Physics and Medicine* 2002; 13(3): 158-63.
- (6) Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine* 2002; 47(6): 1202-10.
- (7) Blaimer M, Breuer F, Mueller M, Heidemann RM, Griswold MA, Jakob PM. SMASH, SENSE, PILS, GRAPPA: How To Choose The Optimal Method. *Top. Magnetic Resonance Imaging* 2004; 15(4): 223-36.
- (8) Wang Y. Description Of Parallel Imaging In MRI Using Multiple Coils. *Magnetic Resonance in Medicine* 2000; 44(3): 495-9.
- (9) Raj A, Zabih R. An Optimal Maximum Likelihood Algorithm For Parallel Imaging In Presence Of Sensitivity Noise. *Proc. ISMRM* 2004: 1771.
- (10) Liang ZP, Bammer R, Ji J, Pelc N, Glover G. Making Better SENSE: Wavelet Denoising, Tikhonov Regularization, And Total Least Squares. *Proc. ISMRM* 2002: 2388.
- (11) Lin F, Kwang K, Belliveau J, Wald L. Parallel Imaging Reconstruction Using Automatic Regularization. *Magnetic Resonance in Medicine* 2004; 51(3): 559-67.
- (12) Bammer R, Auer M, Keeling SL, Augustin M, Stables LA, Prokesch RW, Stollberger R, Moseley ME, Fazekas F. Diffusion Tensor Imaging Using Single-Shot SENSE-EPI. *Magnetic Resonance in Medicine* 2002; 48(1): 128-136.

- (13) Ying L, Xu D, Liang ZP. On Tikhonov Regularization For Image Reconstruction In Parallel MRI. Proc IEEE EMBS 2004: 1056-9.
- (14) Tsao J, Boesiger P, Pruessmann KP. K-T BLAST And K-T SENSE: Dynamic MRI With High Frame Rate Exploiting Spatiotemporal Correlations. Magn Res Medicine 2003; 50(5):1031-42.
- (15) Hansen MS, Kozerke S, Pruessmann KP, Boesiger P, Pedersen EM, Tsao J. On The Influence Of Training Data Quality In K-T BLAST Reconstruction. Magnetic Resonance in Medicine 2004; 52(5):1175--83.
- (16) Tsao J, Kozerke S, Boesiger P, Pruessmann KP. Optimizing Spatiotemporal Sampling For K-T BLAST And K-T SENSE: Application To High-Resolution Real-Time Cardiac Steady-State Free Precession. Magnetic Resonance in Medicine 2003; 53(6): 1372--82.
- (17) Chandra S, Liang ZP, Webb A, Lee H, Morris HD, Lauterbur PC. Application Of Reduced-Encoding Imaging With Generalized-Series Reconstruction (RIGR) In Dynamic MR Imaging. J Magn Reson Imaging 1996; 6(5): 783-97.
- (18) Hanson JM, Liang ZP, Magin RL, Duerk JL, Lauterbur PC. A Comparison Of RIGR And SVD Dynamic Imaging Methods. Magnetic Resonance in Medicine 1997; 38(1): 161-7.
- (19) Kolmogorov V, Zabih R. What Energy Functions Can Be Minimized Via Graph Cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 2004; 26(2): 147-59.
- (20) Boykov Y, Veksler O, Zabih R. Fast Approximate Energy Minimization Via Graph Cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001; 23(11): 1222-39.
- (21) Poggio T, Torre V, Koch C. Computational vision and regularization theory. Nature 1985; 317(6035): 314-9.
- (22) Besag J. On The Statistical Analysis Of Dirty Pictures (With Discussion). Journal of the Royal Statistical Society, Series B 1986; 48(3): 259-302.
- (23) Li S. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.
- (24) Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, And The Bayesian Restoration Of Images. IEEE Transactions in Pattern Analysis and Machine Intelligence 1984; 6(6): 721-41.

- (25) Chellappa R, Jain A. *Markov Random Fields: Theory and Applications*. Academic Press, 1993.
- (26) Cook WJ, Cunningham WH, Pulleyblank WR, Schrijver A. *Combinatorial Optimization*. Wiley, 1998.
- (27) Scharstein D, Szeliski R. A Taxonomy And Evaluation Of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision* 2002; 47(1): 7-42.
- (28) Press W, Teukolsky S, Vetterling W, Flannery B. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.
- (29) Raj A, Zabih R. MAP-SENSE: Maximum a Posteriori Parallel Imaging for Time-Resolved 2D MR Angiography. *Proc. ISMRM 2005*: 1901.
- (30) Greig D, Porteous B, Seheult A. Exact Maximum A Posteriori Estimation for Binary Images. *JRSSB* 1989; 51(2): 271-9.
- (31) Boykov Y, Kolmogorov V. An Experimental Comparison of Min-Cut/Max-Flow Algorithms For Energy Minimization in Vision. *IEEE Transactions on PAMI* 2004; 26(9): 1124-37.
- (32) Kwatra V, Schofield A, Essa I, Turk G, Bobick A. Graphcut Textures: Image and Video Synthesis Using Graph Cuts. *ACM Trans. Graphics, Proc. SIGGRAPH 2003*.
- (33) Boykov Y, Jolly MP. Interactive Organ Segmentation Using graph cuts. *Proc. Medical Image Computing and Computer-Assisted Intervention 2000*: 276-86.
- (34) Kim J, Fisher J, Tsai A, Wible C, Willsky A, Wells W. Incorporating Spatial Priors Into An Information Theoretic Approach For fMRI Data Analysis. *Proc. Medical Image Computing and Computer-Assisted Intervention 2000*: 62-71.
- (35) Raj A, Zabih R. A Graph Cut Algorithm For Generalized Image Deconvolution. *Proc. International Conference of Computer Vision 2005*: 1901.
- (36) Hammer P, Hansen P, Simeone B. Roof duality, complementation and persistency in quadratic 0 - 1 optimization. *Mathematical Programming* 1984; 28: 121-155.
- (37) Reeder SB, Wintersperger BJ, Dietrich O, Lanz T, Greiser A, Reiser MF, Glazer GM, Schoenberg SO. Practical Approaches to the Evaluation of Signal-to-Noise Ratio Performance with Parallel Imaging:

Application with Cardiac Imaging and a 32-Channel Cardiac Coil. *Magnetic Resonance in Medicine* 2005; 54:748–754.

(38) Persson M, Bone D, Elmqvist H. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Phys. Med. Biol* 2001; 46(3): 853-866.

(39) Kolmogorov, Vladimir and Rother, Carsten. Minimizing non-submodular functions with graph cuts - a review. Microsoft Research Technical Report MSR-TR-2006-100, July 2006.

(40) Raj A, Singh G, Zabih R. MRIs for MRFs: Bayesian Reconstruction of MR Images via Graph Cuts. *IEEE Computer Vision and Pattern Recognition Conference* 2006; 1061-1069.

LIST OF TABLES

Table 1: Mean reconstructed SNR and mean g-factor values of reconstructed data for various regularization factors, with $R=3$, $L=8$. Two data sets, Cardiac B and Brain A were used. The means are over object foreground. RSOS mean g value is 1 by definition.

	Cardiac B		Brain A	
	mean SNR	mean g	Mean SNR	Mean g
RSOS	42	1.0	59	1.0
SENSE, $\mu=0.1$	15	3.3	21	3.7
SENSE, $\mu=0.3$	18	2.3	26	2.5
SENSE, $\mu=0.5$	24	1.4	30	1.9
EPIGRAM	36	1.4	40	1.85

Table 2: Quantitative performance measures for reconstruction examples. Both mean reconstructed SNR and mean g-factor values are shown. The means are over object foreground wherever the latter was available. Corresponding reconstructed images were presented in Figs 8-12. Note that numbers for large μ shown below are closer to EPIGRAM numbers, but produce images with considerably more aliasing.

	R	SENSE, optimal μ		SENSE, large μ		EPIGRAM	
		mean SNR	Mean g	Mean SNR	Mean g	Mean SNR	Mean g
Brain A	4	8	4.6	13	2.4	23	1.7
Brain B	5	10	3.5	13	2.4	17	2.2
Cardiac A	3	20	2.3	25	1.6	33	1.5
Cardiac B	3	15	3.3	18	2.3	36	1.4

LIST OF FIGURES

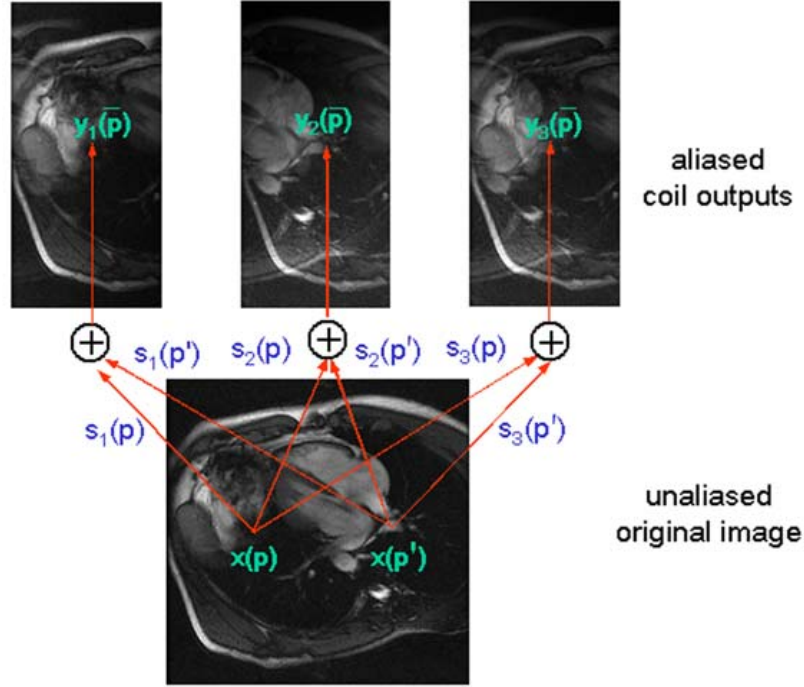


Figure 1: A schematic of the pixel-wise aliasing process, for a single pair of aliasing pixels p and p' , for 2x acceleration using 3 coils. The aliased observations Y_1 are obtained by a weighted sum of the aliasing pixels, weighted by coil sensitivity values S_1 . To simplify the figure, aliasing is shown in the horizontal direction.

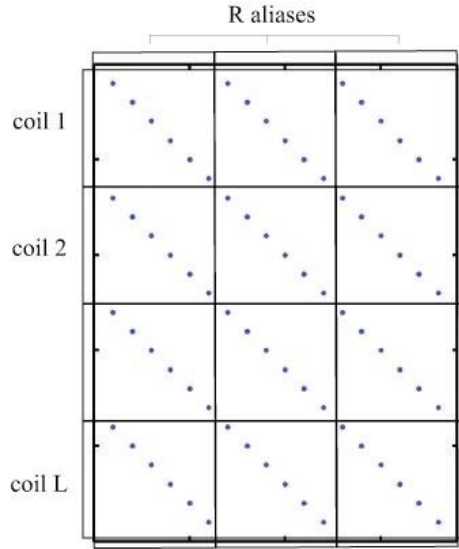


Figure 2: The $L \times R$ block-diagonal structure of matrix \mathbf{E} . Each sub-block sensitivity-encodes one aliasing band in the image. Matrix \mathbf{E} is a concatenation over all coil outputs, as shown by coil labels.

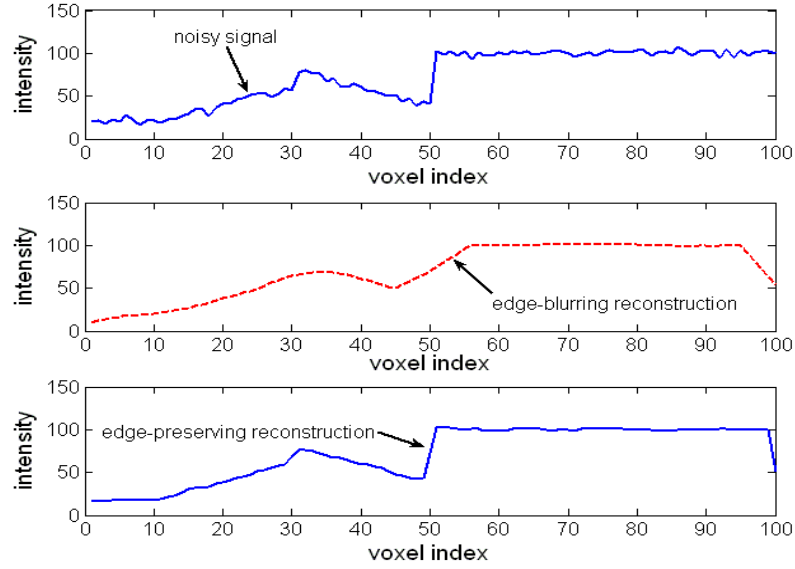


Figure 3: Typical performance of edge-preserving and edge-blurring reconstruction on noisy image row. Fig (b) was obtained by globally smoothing via a Gaussian kernel, Fig (c) by an edge-preserving median filter.

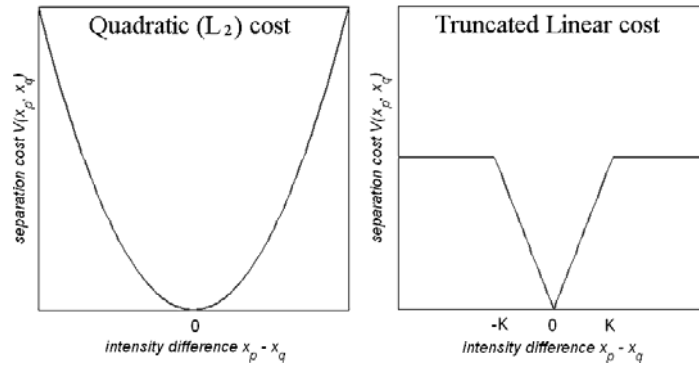


Figure 4: Two natural separation cost functions for spatial priors. The L_2 cost on the left usually causes edge blurring due to excessive penalty for high intensity differences, whereas the truncated linear potential on the right is considered to be edge-preserving and robust. For MR data, the truncated linear model appears to work best. While the L_2 separation cost does not preserve edges, its convex nature vastly simplifies the optimization problem.

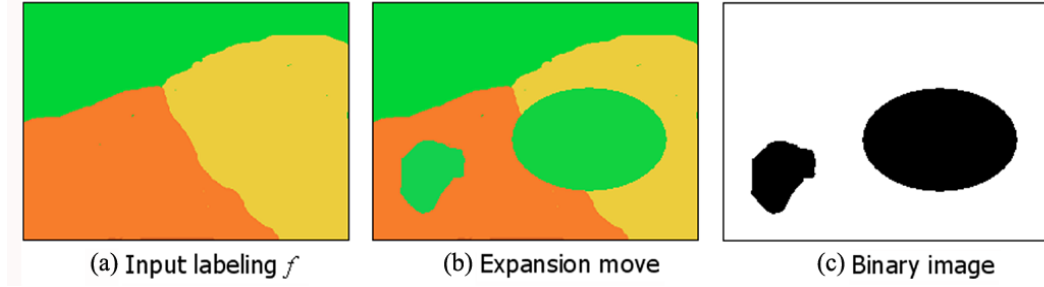


Figure 5: The expansion move algorithm. Start with the initial labeling of the image shown by different colours in (a), assuming only 3 labels. Note that here labels correspond to intensities, although this need not be the case in general. First find the expansion move on the label “green” that most decreases objective function \mathcal{E} , as shown in (b). Move there, then find the best “orange” expansion move, etc. Done when no α -expansion move decreases the cost, for any label α . Corresponding to every expansion move (b) is a binary image (c), where each pixel is assigned 1 if its label changes, and 0 if it does not.

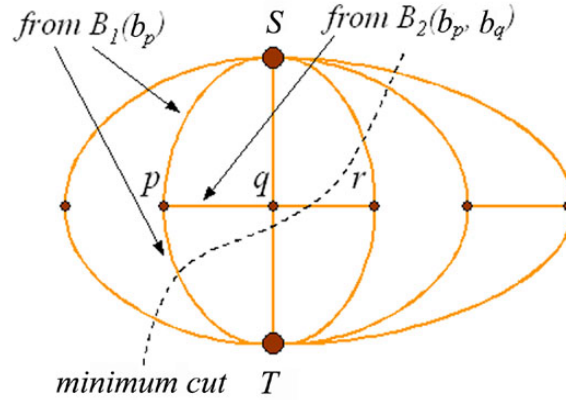


Figure 6: Graph construction for minimizing a given binary objective function. The graph consists of nodes corresponding to image voxels, as well as two special “terminal” nodes ‘S’ and ‘T’. Edges between nodes represent single-voxel (B_1) and pairwise cost (B_2) terms. A minimum cut on this graph solves the binary minimization problem.

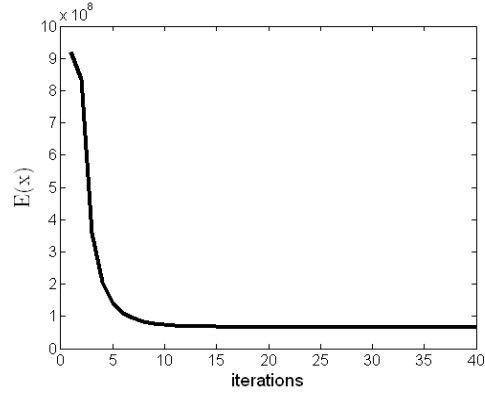


Figure 7: Convergence behaviour of the modified graph cut algorithm (EPIGRAM) for acceleration factor $R = 3$, $L = 8$ coils, on GRE torso data. The vertical axis shows the value of objective function [15] achieved after each outer iteration, which represents a single cycle through $N_{\text{labels}} = 256$ expansion moves.

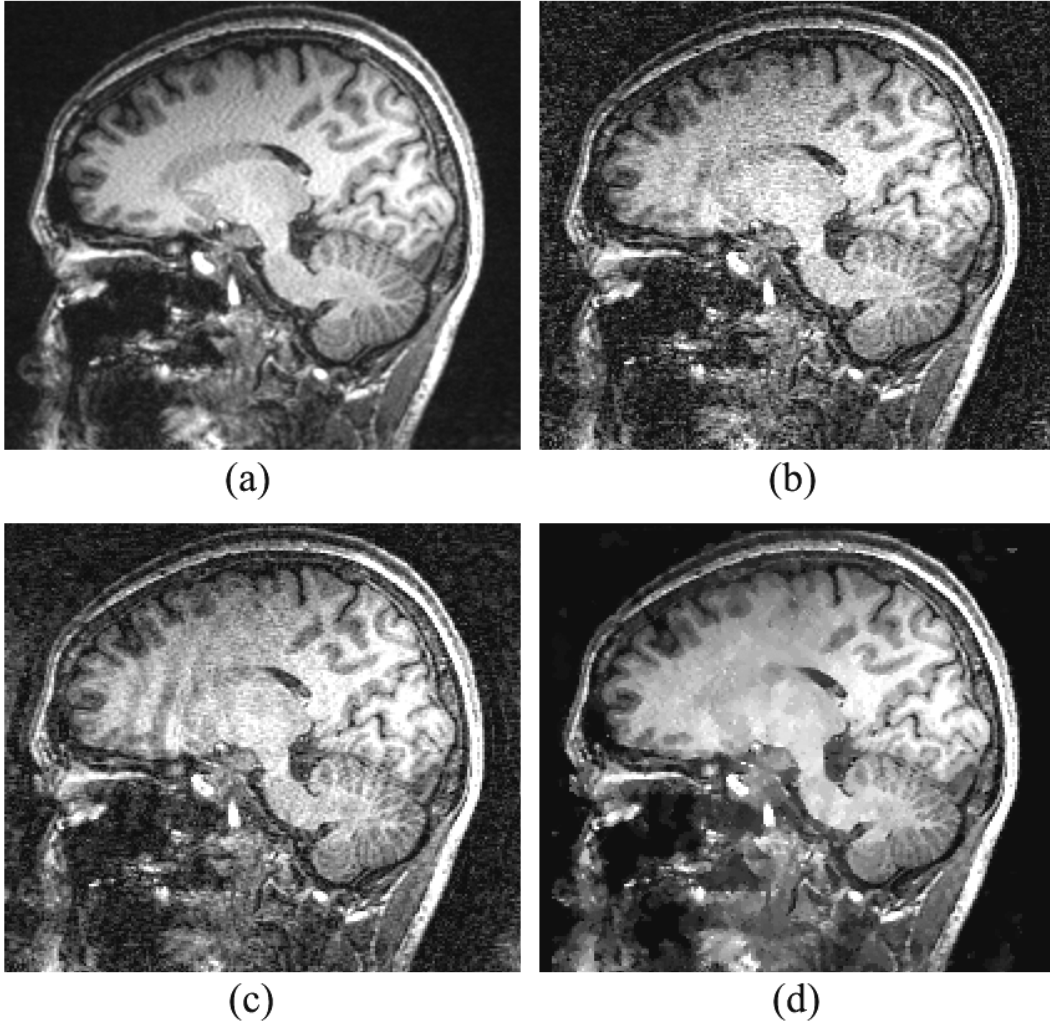


Figure 8: Brain A: In vivo brain result with $R = 4$, $L=8$. Views were acquired vertically. (a) Reference image (b) SENSE regularized with $\mu = 0.08$ (c) SENSE regularized with $\mu = 0.16$, (d) EPIGRAM reconstruction

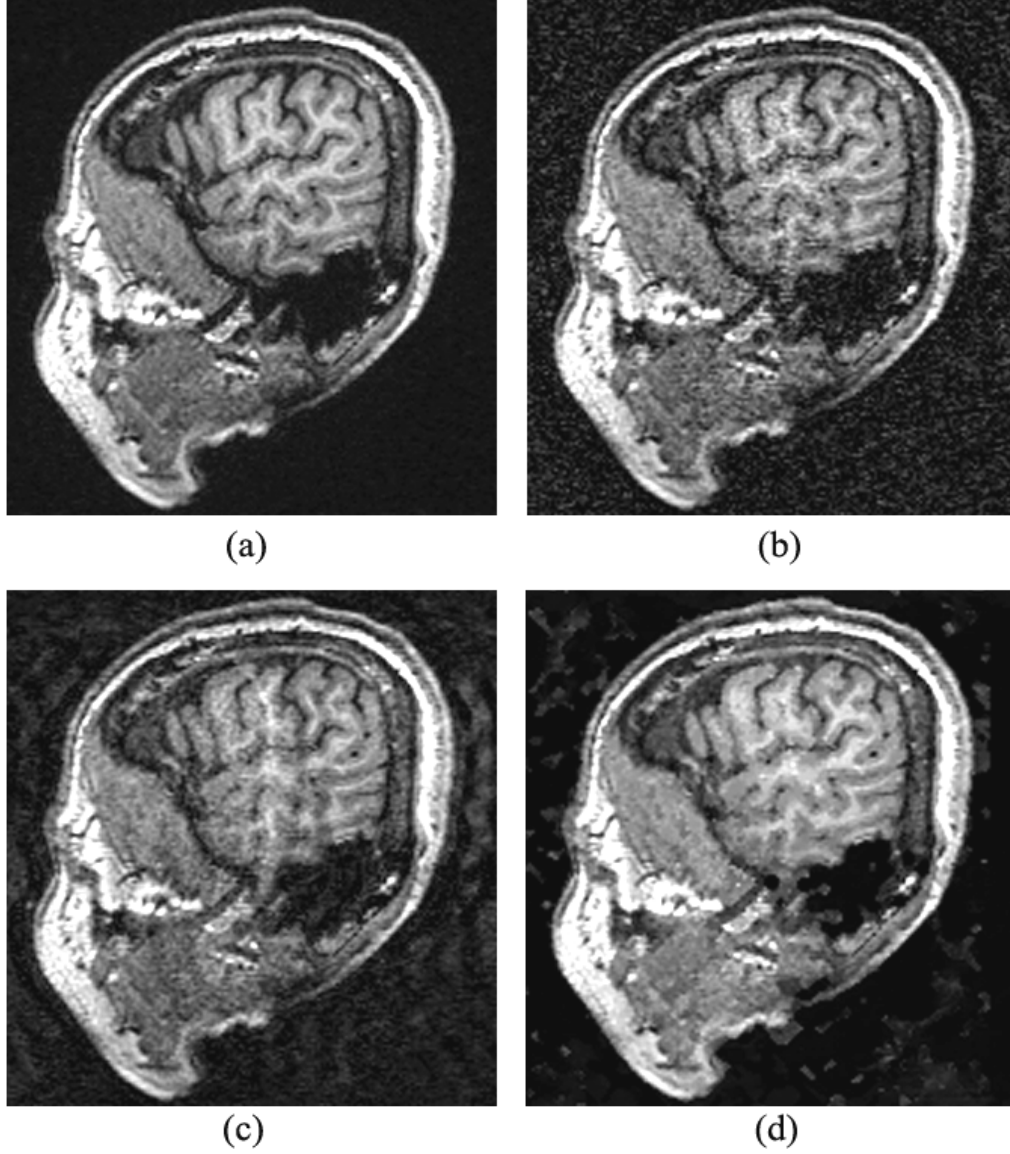


Figure 9: Brain B: In vivo brain result with $R = 5$, $L=8$. Views were acquired vertically. (a) Reference image (b) SENSE regularized with $\mu = 0.1$ (c) SENSE regularized with $\mu = 0.2$, (d) EPIGRAM reconstruction

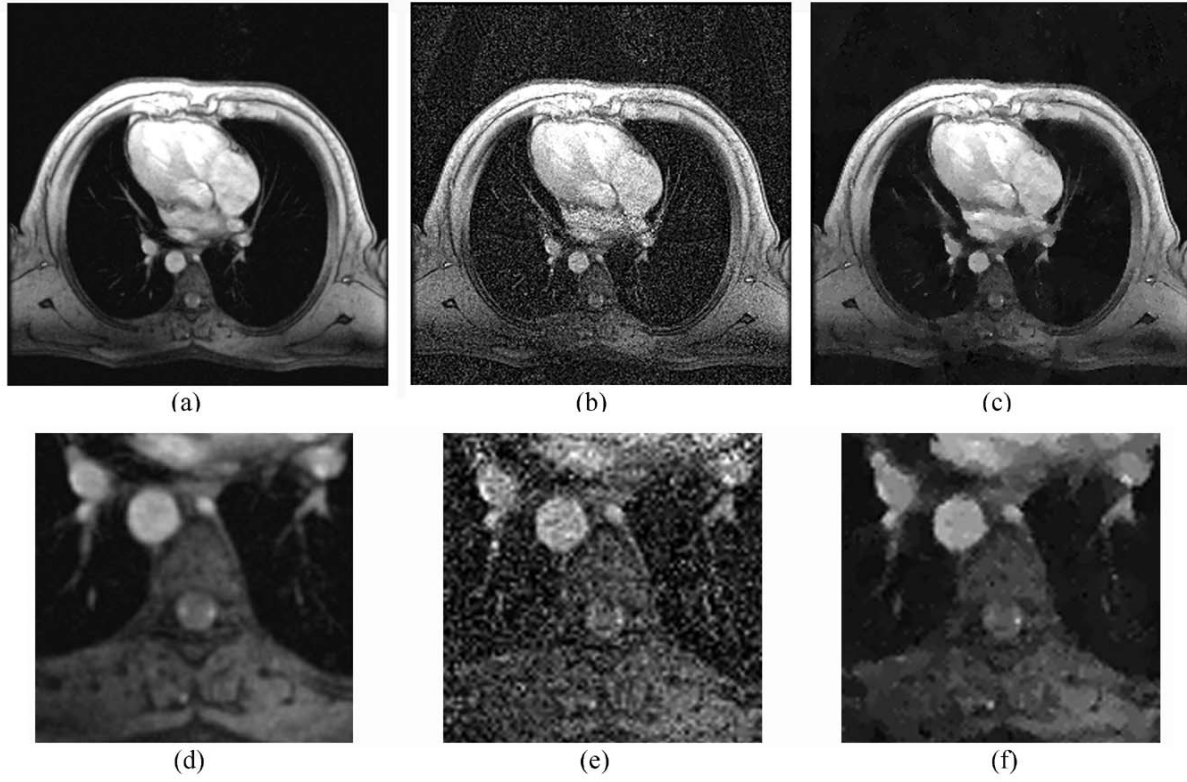


Figure 10: Cardiac A: *in vivo* result $R = 3$, $L=8$. Views were acquired horizontally. (a) Reference image (b) SENSE regularized with $\mu = 0.16$ (optimal) (c) EPIGRAM reconstruction. Zoomed-in portions of (a)-(c) are shown in (d)-(f).

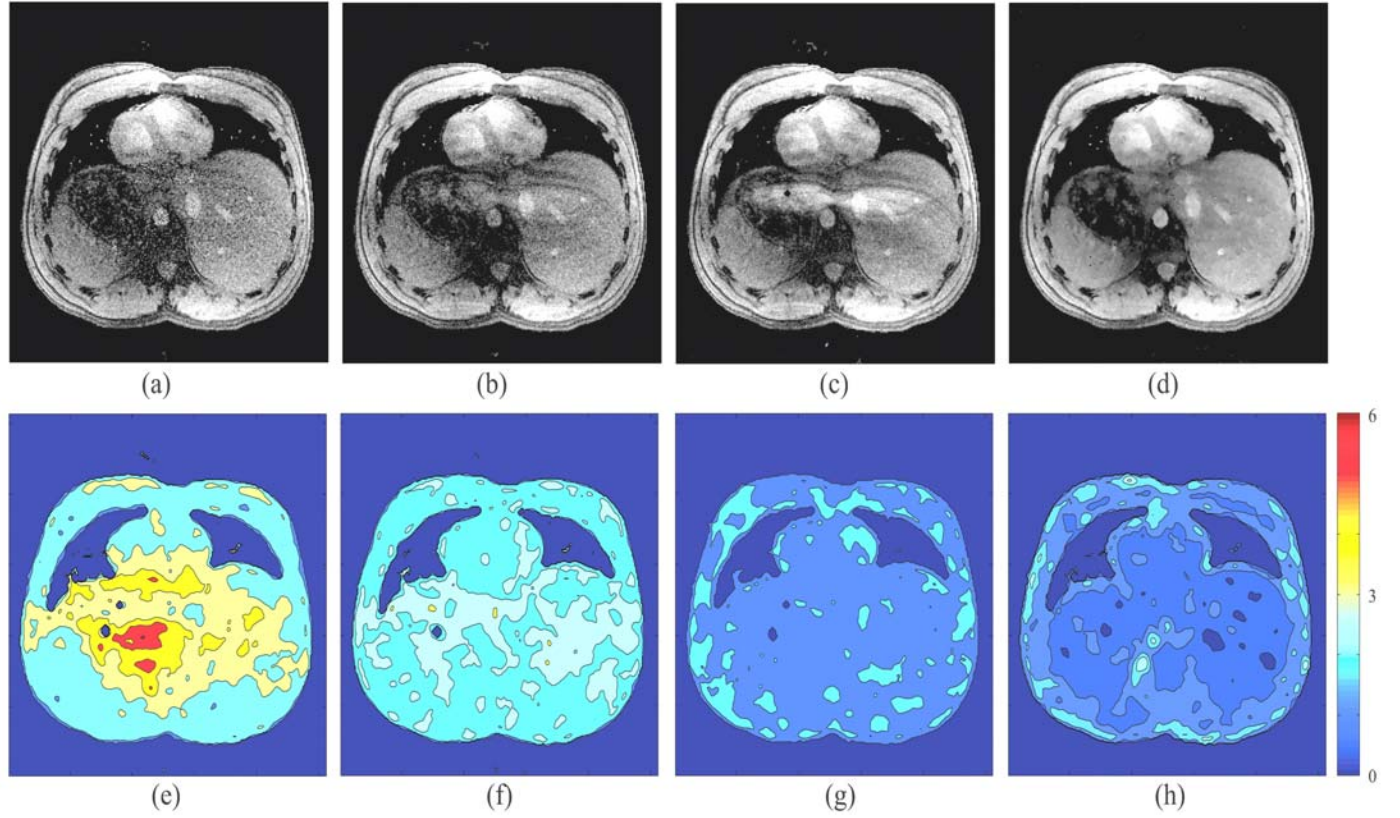


Figure 11: Cardiac B: *in vivo* result $R = 3$, $L=8$ - effect of regularization on g-factor maps. Views were acquired horizontally. Figures (a) and (e) show SENSE reconstruction and its g-map for $\mu = 0.1$. There is inadequate noise suppression, with a g-factor as high as 6. Results for $\mu = 0.3$ are shown in (b), (f). The g-maps has become correspondingly flatter, but at the cost of aliasing. The rightmost column (d), (h) shows EPIGRAM results, which has significantly lower g values as well as lower aliasing artifacts. SENSE with $\mu = 0.5$ (c), (g) matches the EPIGRAM g-map, but yields unacceptable aliasing.

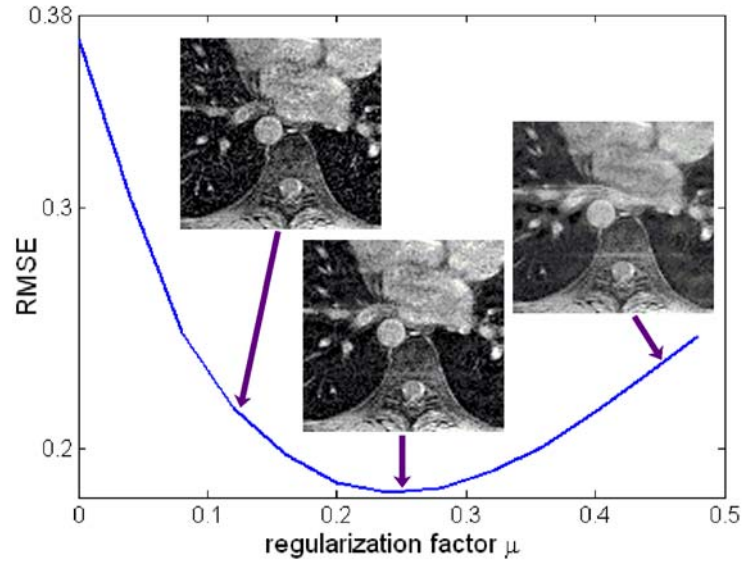


Figure 12: Average RMSE of regularized SENSE reconstruction of torso data. The optimum was observed at around 0.25, although visual quality was found to be best for smaller values, in the region [0.1-0.2]. Three example torso reconstructions are also shown. Observe that as regularization increases, residual aliasing artifacts and blurring both get worse. Also note that non-regularized SENSE (i.e., $\mu = 0$) gives substantially worse RMSE.

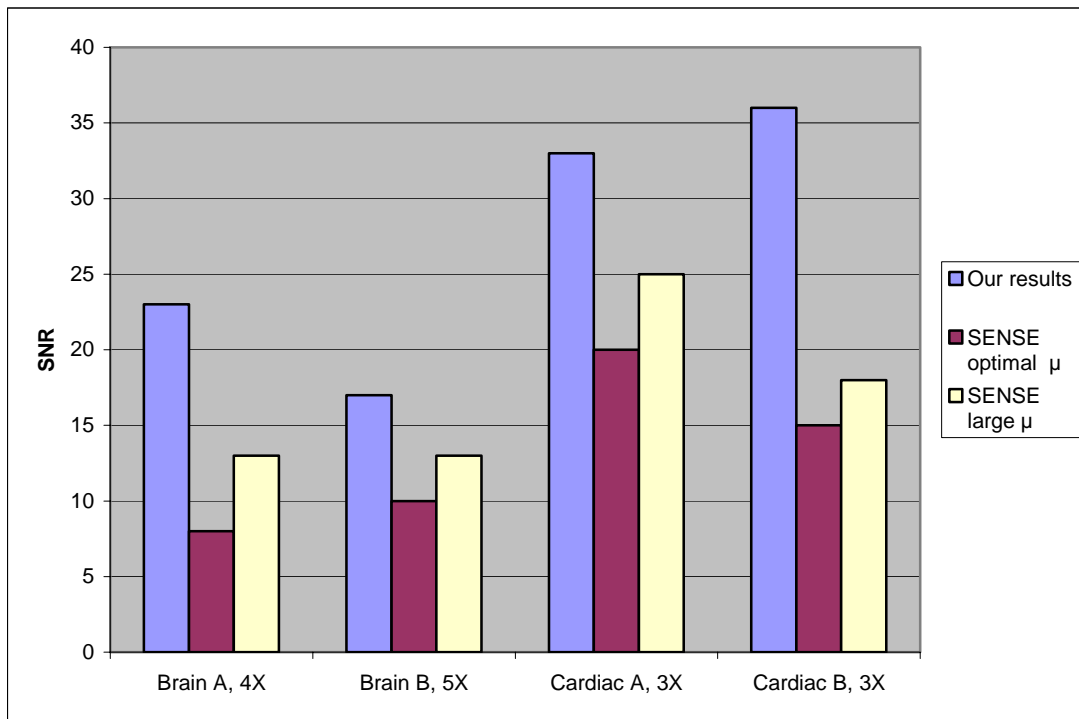


Figure 13: A plot of SNR achieved by EPIGRAM and SENSE for various data sets. SNR values were taken from table 2. Observe that large μ gives higher SNR, but from the displayed images, this comes at the cost of unacceptable aliasing.