UCSF

UC San Francisco Previously Published Works

Title

A dictionary-based graph-cut algorithm for MRI reconstruction

Permalink

https://escholarship.org/uc/item/2421f98n

Journal

NMR in Biomedicine, 33(12)

ISSN

0952-3480

Authors

Xu, Jiexun

Pannetier, Nicolas

Raj, Ashish

Publication Date

2020-12-01

DOI

10.1002/nbm.4344

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

NMR Biomed. Author manuscript; available in PMC 2022 June 03.

Published in final edited form as:

NMR Biomed. 2020 December ; 33(12): e4344. doi:10.1002/nbm.4344.

A Dictionary-based Graphcut Algorithm for MRI Reconstruction

Jiexun Xu[1], **Nicolas Pannetier**[2], **Ashish Raj**[3]

[1]Department of Computer Science, Cornell University, Ithaca, NY, USA

[2]Department of Radiology, University of California, San Francisco, CA, USA

[3]Department of Radiology, University of California, San Francisco, CA, USA

Abstract

Purpose: Compressive sensing based image reconstruction methods have proposed random under-sampling schemes that produces incoherent, noise-like aliasing artifacts, which are easier to remove. The denoising process is critically assisted by imposing sparsity-enforcing priors. Sparsity is known to be induced if the prior is of the form of $Lp\,(0 \le p \le 1)$ norm. CS methods generally use a convex relaxation of these priors like the L_1 norm, which may not exploit the full power of CS. An efficient, discrete optimization formulation is proposed which works not only on arbitrary Lp norm priors as some non-convex CS methods do, but also on highly non-convex truncated penalty functions, resulting in a specific type of edge preserving priors. These advanced features make the minimization problem highly non-convex, and thus call for more sophisticated minimization routines.

Theory and Methods: The work combines edge-preserving priors with random under-sampling, and solve the resulting optimization using a set of discrete optimization methods called Graph Cuts. The resulting optimization problem is solved by applying graph cuts iteratively within a dictionary, defined here as an appropriately constructed set of vectors relevant to brain MRI data used here.

Results: Experimental results with in vivo data are presented.

Conclusion: The proposed algorithm produces better results than regularized SENSE or standard compressive sensing for reconstruction of in vivo data.

Keywords

Parallel Imaging; SENSE; Compressive Sensing; Graph Cuts

INTRODUCTION

Undersampled MRI acquisitions can speed up MRI scan time, but it also introduces aliasing artifacts due to sub-Nyquist sampling in k-space, if reconstructed using a simple inverse FFT of zero-filled k-space data. Aliasing is removed during reconstruction using advanced methods like parallel imaging, which uses the redundancy provided by multiple receiver coils, and by Compressive Sensing (CS), which exploits the noise-like aliasing artifacts

Corresponding Author: jiexun.xu@gmail.com.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

resulting from random acquisition schemes. Traditional parallel imaging reconstruction algorithms include Sensitivity Encoding (SENSE) (1–3), simultaneous acquisition of spatial harmonics (SMASH) (4), (5), and generalized auto-calibrating partially parallel acquisitions (GRAPPA) (6). Although these methods can be adapted for the case of arbitrary under-sampling patterns, their usual implementation is for uniform Cartesian under-sampling. Such under-sampling results in *structured* aliasing artifacts called image folding, whereby many shifted copies of the true image are folded on top of each other. Uniform Cartesian parallel imaging methods can allow for acceleration of up to 2–4 times, but further acceleration is limited by the ill-conditioning of the inverse problem (7). Although numerical conditioning is frequently improved by Tikhonov regularization schemes (8), (9), this reduces the effectiveness of anti-aliasing, and produces unrealistically smooth images. In contrast, CS generally relies on *unstructured* or *incoherent* artifacts resulting from randomized k-space under-sampling schemes (10–12), which are usually removed by imposing sparsity-inducing prior terms within an energy minimization framework, commonly via the L_1 norm of the desired image (such as angiograms) or of a certain sparsifying transform, such as the first difference transform or the wavelet transform (11–13).

In this paper, we present, for the first time, a discrete optimization algorithm for effectively combining PI and CS approaches based on *graphcuts* (14), a well-established algorithm for solving binary optimization problems. Although some previous work on combining CS and PI has been reported (15), (16), the motivation of the current work is that current CS or PI methods cannot effectively impose arbitrary non-convex penalty functions. Indeed, it is well-understood that in general, convex L_1 penalties used in CS (such as lasso (17)) are not the most appropriate means of imposing sparsity measured by the L_0 norm. Although CS theory proves that under certain conditions, solving the L_1 problem also solves the L_0 problem (18), such conditions are not consistently met in MRI. For instance, a strict requirement is that the linear system given by the matrix E must have strictly uncorrelated rows. Even under random undersampling in MRI, usually the center of k-space is more densely sampled than its periphery in order to take advantage of higher SNR in the center. However, this usually violates the strict non-correlation condition. Therefore, most CS algorithms cannot directly minimize these truly sparsifying functions; instead, much work has focused on identifying conditions under which sparsity may be obtained from convex approximations (19).

Although this approach has proved successful, further progress in the field will require accommodation of truly non-convex functions. Specifically, sophisticated non-convex “edge preserving priors” (EPP) involving sub-linear norms with truncated penalties have been shown to be very powerful in low-level computer vision applications (20–22), statistics (23), and image processing (24). Compared to conventional Tikhonov regularization, EPPs are a general class of spatial smoothness priors that do not indiscriminately blur tissue boundaries. These non-convex and non-smooth EPPs are natural models of robust image statistics, and have a close relationship with outlier resistance (25–27). While these priors have a strong tradition in computer vision, their use in medical imaging is scarce, such as brain MRI (28) and cardiac MRI (29). Certain CS algorithms, such as (30–33), can handle limited non-convexity, but not the kind of EPP regularization terms that are our specific focus here – those that allow for sub-linear or truncated behavior at high edge-related intensity gradients. These algorithms are also not natively non-convex, but instead rely on successively closer

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

convex approximations or greedy approaches, causing them to be susceptible to convergence to weak or inappropriate local minima.

In contrast, graphcut routines (14), (34), (35) were developed to solve the challenging non-convex minimization problems, and have found success in applications such as stereo matching (36), segmentation (37) and reconstruction (38). Although graphcut routines cannot provide global optimal *discrete/quantized* solution in general (39), they can converge to "strong local minima", the best solution within a large radius (40). These strong properties motivate us to apply graphcut techniques to the combined PI+CS problem with the complicated EPP prior. Our proposed graphcut algorithm can handle much more general non-convex functions in practically feasible compute times, and works for arbitrary k-space sampling schemes and arbitrary number of parallel receiver coils. Our work combines, for the first time, the strengths of graphcuts (*via* edge preserving priors, ability to solve non-convex problems, and fast convergence to strong local minima) with those of CS (*via* arbitrary sampling schemes, incoherent aliasing, and sparsity prior in transform domains). Our algorithm performs reconstruction on the entire volumetric data, so it can take advantage of true 3D reconstructions.

Our work is related to previous graphcut algorithm called "Edge-preserving Parallel Imaging with Graph-cut Minimization" (EPIGRAM) (28), which deployed EPPs for improving the reconstruction of *uniform* Cartesian undersampled parallel imaging data and proved that such priors are suitable models for MR images. Uniform Cartesian under-sampling yields a unique sparse block-diagonal matrix structure, whose features were exploited by EPIGRAM to obtain a sparse graph on which graphcut operations are very efficient. A variant of EPIGRAM (29) has also appeared that speeds up the EPIGRAM process using *jump moves* (41). Although both EPIGRAM and the current proposal use graphcuts, moving from uniform Cartesian to arbitrary sampling trajectories necessitates a very different formulation, because the graph is no longer sparse, and a direct application of EPIGRAM would be computationally prohibitive.

Our algorithm works as follows. The PI+CS problem is posed as the minimization of an energy function imposing a data fidelity term and an EPP term. Since this energy function cannot be directly minimized due to non-convexity, we reformulate it as a series of binary minimization problems. We start from an initial image obtained from regularized SENSE. Each binary optimization problem is posed as a *dictionary search*, such that each member of a pre-constructed dictionary has a chance of contributing to the desired image – a binary choice for each member. We show that this dictionary search constitutes a binary optimization problem efficiently solvable with graphcut techniques called "Quadratic Pseudo-Boolean Optimization" (QPBO) (35). Dictionary based methods have been used in learning (42) and image denoising (43), where a large number of patches from many artifact-free pristine images are treated as members of the dictionary. In this work, we obtain a dictionary of high frequency image features like edges, but unlike other methods, our dictionary comes directly from the undersampled dataset at hand. Our dictionary search strategy greatly reduces the number of binary moves necessary for convergence, and dramatically reduces the size and complexity of each binary problem. We show that the resulting algorithm can efficiently solve the combined CS+PI problem, even for highly non-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

convex penalty functions. We demonstrate that our algorithm produces *in vivo* results with higher numerical and visual acuity compared to regularized SENSE and CS. Subsequent brain volumetric analysis produced lower error rates compared to other methods. We believe this is the first graphcut algorithm for solving arbitrary non-convex problems involving random sampling schemes. Although the current application is for accelerated MRI, due to its generality, we believe this algorithm has important applicability to a host of other imaging problems.

THEORY

Parallel Imaging Summary

Suppose the MRI machine has c coils that sample, in Cartesian space, a m-by-n-by-s image. Let $M=mns$. Then the joint PI/CS problem can be written as a discretized linear system

$$\boldsymbol{y} = GFS\boldsymbol{x} + noise \quad [1]$$

$\boldsymbol{x}$ is the vectorization of the target m-by-n-by-s image. S is a Mc-by-M block matrix with diagonal blocks that encode the sensitivity responses from all coils. F is a Mc-by-Mc block diagonal matrix where each block is a discretized 2D Fourier transform matrix. G is Mc-by-Mc 0–1 diagonal matrix that encodes the under-sampling scheme defined on a Cartesian grid. Thus, each voxel in $\boldsymbol{x}$ is first multiplied by sensitivity response of each coil, which then undergoes a 2D Fourier transform into k-space, giving $FS\boldsymbol{x}$. Finally the signals are sampled with G to obtain the phased array signal $GFS\boldsymbol{x}$, or $\boldsymbol{y}$. Fig 1 gives a visual demonstration of the structure of these matrices.

Let $E = GFS$. Eq.1 simplifies to $E\boldsymbol{x} = \boldsymbol{y}$. SENSE reconstructs the image via the pseudo-inverse $\boldsymbol{x}_{SENSE} = (E^H E)^{-1} E^H y$. Unfortunately, inverse problems of this form become progressively ill-posed with increasing acceleration, leading to noise amplification and insufficient anti-aliasing.

Compressive sensing methods exploits the sparsity of the solution, and have demonstrated promising results (11), (13), (15), (16). In particular, (13) has proposed an objective with both total variation penalizers and sparsity penalizers in the wavelet domain:

$$\arg\min_{\boldsymbol{x}} \|E\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda_1 TV(\boldsymbol{x}) + \lambda_2 \|W\boldsymbol{x}\|_1 \quad [2]$$

Instead of TV and sparsity penalizers, a different class of edge-preserving priors (EPP) assumes piecewise-smoothness of medical images. In particular, EPIGRAM introduces a truncated L_p-norm prior in the image domain, which leads to the following objective:

$$\arg\min_{\boldsymbol{x}} \|E\boldsymbol{x} - \boldsymbol{y}\|^2 + G_{EP}(\boldsymbol{x}) \quad [3]$$

The EPP prior is $G_{EP}(\boldsymbol{x}) = \lambda \sum_{i,j} \min\left(|x_i - x_j|^p, T\right)$, where T and p are constant. This is the same non-convex prior used in EPIGRAM. The idea behind this prior lies in penalizing

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

non-smoothness of neighboring voxels, but the introduction of the truncation parameter T prevents over-penalization of discontinuous voxel intensity across image edges.

Graphcut and QPBO

Unfortunately, Eq.3 is highly non-convex. Graphcut algorithms address this problem by converting the original energy minimization into a sequence of binary minimization problems. The EPIGRAM (28) method used binary problems arising from the so-called *expansion moves*. Here we employ more general moves of the form:

$$\arg\min_{\boldsymbol{b}^k} \left\| E\left(\boldsymbol{x}^{k-1} + U^k \boldsymbol{b}^k\right) - \boldsymbol{y} \right\|^2 + G_{EP}\left(\left(\boldsymbol{x}^{k-1} + U^k \boldsymbol{b}^k\right)\right) \quad [4]$$

where $\boldsymbol{x}^{k-1}$ is the solution of (3) at (k-1)-th iteration, U^k is a matrix, and $\boldsymbol{b}^k = \{\mathbf{0}, \mathbf{1}\}^{\boldsymbol{n}}$ is an unknown binary vector, to be solved at iteration k. If U^k is a M-by-M diagonal matrix with $U^k_{i,\,i} = \alpha^k - \boldsymbol{x}_i^{k-1}$ for some constant α^k, then Eq.4 is equivalent to the α-expansion move variant used in EPIGRAM, where each binary variable can choose to stay at $\boldsymbol{x}_i^{k-1}$ or move to α^k. If U^k is a M-by-M diagonal matrix with equal diagonal entries β^k, then Eq.4 is equivalent to *jump moves*, where each binary variable can choose to stay at $\boldsymbol{x}_i^{k-1}$ or jump to $\boldsymbol{x}_i^{k-1} + \beta^k$. If U^k is an arbitrary M-by-M diagonal matrix, then Eq.4 is equivalent to *fusion moves* (44) where each binary variable can choose to stay at $\boldsymbol{x}_i^{k-1}$ or move to $\boldsymbol{x}_i^{k-1} + U^k_{i,\,i}$. Eq.4 was expressed as a summation of unary and pairwise terms of unknown variables $\boldsymbol{b}^k$. A sparse weighted graph was then constructed, whose vertices represented unknown variables plus two special "source" and "sink" vertices, and whose weighted edges represented the unary and pairwise terms, as described in (28) and (29). Once the graph was constructed, a *minimum-cut* (45) on this graph was generated, which provably provided the globally optimal solution to Eq.4 when the energy function satisfy the sub-modularity property (39). Of course, optimal solution of successive expansion or jump moves would not normally guarantee optimal solution of the original energy function, yet it is well documented that successive jump moves do generate strong local minima for many image-level applications (46), (47). Traditional graphcut algorithm can only handle a certain class of non-convex functions called submodular functions (39), but recent advances have introduced a new class of algorithm called Quadratic Pseudo-Boolean Optimization (QPBO) (34), which expands the graphcut algorithm to handle non-submodular functions. While QPBO is not guaranteed to find global optimum for non-submodular functions, it performs well in practice.

The major limitation of EPIGRAM, whether using expansion, jump or fusion moves, is the requirement of efficiently constructing a graph – one which cannot be met for random under-sampling patterns used in CS. In this work, we replace the diagonal matrix U^k with a dictionary, a matrix with more rows than columns. This greatly reduces the dimensionality of $\boldsymbol{b}^k$ and the problem size, allowing graphcut to be applied to the resulting problem.

Algorithm Overview

During the minimization of [4], E^*E is computed. With Cartesian under-sampling, this matrix is sparse due to the special way Cartesian sampling was done (28). Thus, the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

EPIGRAM technique can be applied. However, with random under-sampling, E^*E is dense, so the number of pairwise terms is quadratic in the number of variables. E^*E is dense because edges of the graph are related to entries of the system matrix E, which in turn is sparse under regular undersampling but dense under random undersampling. This precludes EPIGRAM from being applied directly due to the prohibitive cost of constructing a dense graph with millions of vertices. To overcome this constraint, we introduce a dictionary V_0 and an initial solution $\boldsymbol{x}^0$ into Eq.3 and reformulate it as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\| E\left(\boldsymbol{x}^0 + V_0\boldsymbol{\alpha}\right) - \boldsymbol{y} \right\|^2 + G_{EP}\left(\left(\boldsymbol{x}^0 + V_0\boldsymbol{\alpha}\right)\right) \quad [5]$$

$$\hat{\boldsymbol{x}} = \boldsymbol{x}^0 + V_0\hat{\boldsymbol{\alpha}}$$

where $\hat{\boldsymbol{x}}$ is the final solution and $\boldsymbol{\alpha}$ is an unknown vector of coefficients which, if it could be efficiently found, would provide the lowest energy weighted combination of the dictionary members. Despite a significant reduction of problem size due to V_0, the minimization problem Eq.5 is still non-convex and difficult to solve. To overcome these challenges, we rely on a series of smaller and more plausible dictionaries V^k, chosen with a heuristic in step 5a below, whose members are mutually disjoint (the support of each member does not overlap), and perform a series of fast binary minimizations to solve Eq.5. We note that our use of the term “dictionary” is not entirely consistent with its usage in dictionary-based learning; nonetheless we use it here due to the intuition that feature vectors may be precomputed from available data that may be collected in a dictionary of features.

Denote $|\boldsymbol{x}|$ as the image obtained by taking the absolute value of each component of the complex image $\boldsymbol{x}$. Our proposed algorithm is sketched below (details are discussed in the next section):

1. Compute an initial solution $\boldsymbol{x}^0$ with regularized SENSE by minimizing the following problem:

$$\boldsymbol{x}^0 = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|E\boldsymbol{x} - \boldsymbol{y}\|^2 + \|\boldsymbol{x}\|_2^2$$

2. Use $|\boldsymbol{x}^0|$ to construct image segmentation $S = \{\boldsymbol{s}^j, j = 1, \ldots, J\}$ consisting of J segments where each $\boldsymbol{s}^j$ is a binary vector denoting the jth segment
3. Construct a master dictionary V_0 using $\boldsymbol{x}^0$ and segmentation S
4. Precompute the matrix $M^0 = V_0^* E^* E V_0$ (In this paper, A^* denotes conjugate transpose of matrix A). Note that the dimension of M^0 equals the total number of columns in V_0, the total number of dictionary vectors. By only choosing a moderate number of dictionary vectors M^0 is not prohibitively expensive to be computed and used in following steps, and this precomputation will not be very time consuming.
5. Iterate over k=1, 2, …:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a. Minimize the following simple, convex problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\| E\left(\boldsymbol{x}^{k-1} + V_0\boldsymbol{\alpha}\right) - \boldsymbol{y} \right\|^2 + \lambda \left\| D\left(\boldsymbol{x}^{k-1} + V_0\boldsymbol{\alpha}\right) \right\|^2 \quad [6]$$

where λ is a regularization parameter and D is a first difference operator which penalizes the intensity difference between neighbouring voxels

b. Use $\hat{\boldsymbol{\alpha}}$ and the segments $j = 1, \ldots, J$ to construct a mutually disjoint dictionary $V^k = \left[\boldsymbol{v}_1^k, \ldots, \boldsymbol{v}_J^k\right]$, such that $\boldsymbol{v}_j^k = \boldsymbol{s}^j \cdot (V_0\tilde{\boldsymbol{\alpha}})$ (element-wise multiplication)

c. Solve the following binary minimization problem using QPBO

$$\hat{\boldsymbol{b}}^k = \arg\min_b \left\| E\left(\boldsymbol{x}^{k-1} + V^k\boldsymbol{b}\right) - \boldsymbol{y} \right\|^2 + G_{EP}\left(\left(\boldsymbol{x}^{k-1} + V^k\boldsymbol{b}\right)\right) \quad [7]$$

Where $\hat{\boldsymbol{b}}^k$ is a binary vector whose values are either 0 or 1

d. Update $\boldsymbol{x}^k = \boldsymbol{x}^{k-1} + V^k\hat{\boldsymbol{b}}^k$

6. Terminate when converged, and output $\hat{\boldsymbol{x}} = \boldsymbol{x}^k$. We consider the algorithm converged if $\hat{\boldsymbol{b}}^k$ is all zero (no moves are committed), or a maximum number of iterations have passed

We begin with an initial image and its segmentation S, which is used to tile the image and dictionary members into a set of J disjoint segments. We precompute $M^0 = V_0^* E^* E V_0$ to accelerate step 5c of our algorithm. We use the solution to Eq.6 to construct a mutually disjoint dictionary V_k that stores move proposals for each segment. A non-convex binary minimization problem is formulated using V_k to determine whether these plausible move proposals are committed or not – a problem efficiently solved by QPBO. It is worth noting that if the move proposals in V^k are implausible (i.e. they result in a higher energy if committed), then most of $\hat{\boldsymbol{b}}^k$ will be zero and $\mathbf{x}^k$ will be similar to $\mathbf{x}^{k-1}$. Define the ratio of number of ones in $\hat{\boldsymbol{b}}^k$ to its dimension as PAR, Proposal Acceptance Ratio. The minimization in Eq.6 helps to quickly generate plausible move proposals, which lead to high PAR to meaningfully evolve $\mathbf{x}^k$ at each iteration. Note that upon convergence, $\hat{\boldsymbol{x}} = \boldsymbol{x}^0 + V_0\hat{\boldsymbol{\alpha}} = \boldsymbol{x}^0 + \sum_{k=1}^{N} V^k\hat{\boldsymbol{b}}^k$. $\hat{\boldsymbol{\alpha}}$ can thus be interpreted as combined contributions from each dictionary vector. Such contribution is iteratively summed up from the binary vectors $\hat{\boldsymbol{b}}^k$ and their dictionaries V^k

In summary, our algorithm can be viewed as fusion moves with segments of the image as unknown variables instead of individual voxels.

The mutually disjoint property of dictionary V_k is necessary for computational reasons. Recall that $G_{EP}\left(\left(\boldsymbol{x}^{k-1} + V^k\boldsymbol{b}\right)\right) = \lambda \sum_{i,j} \min\left(\left|\left(x_i + \sum_q V_{i,q}^k b_q\right) - \left(x_j + \sum_r V_{j,r}^k b_r\right)\right|^p, T\right)$, where x_i and x_j are the i-th and j-th intensity value of $\mathbf{x}^{k-1}$, and $V_{i,q}^k$ and $V_{j,r}^k$ are the *(i,q)*-the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and (j,r)-th element of V^k, respectively. To minimize Eq.7 with QPBO, all possible energy contributions need to be computed from each summation term. If the i-th and j-the rows of V^k have m and n non-zero entries respectively, then $m+n$ binary variables will be involved for the pair of voxels (x_i, x_j) in the smoothness term, which leads to 2^{m+n} possible energy terms. Although QPBO can be applied to energy functions whose terms have arbitrarily large number of binary variables, in practice they work best for cases where there are no more than 2 or 3 binary variables in any term. The above mutually disjoint design of V^k limits the number of binary variables per term to at most 2, such that all summation terms in the binary energy function are either unary or pairwise. This is necessary for efficiently constructing the graph that represents Eq.7.

Details of implementation

Segmentation.—We use (48) for segmentation. The algorithm defines a predicate for measuring evidence of boundary with a graph-based representation of the image, and segments the image based on this predicate. The algorithm is modified slightly to handle 3D data. Although the purpose of obtaining mutually disjoint dictionary vectors could be met by an arbitrary tiling of the image, we chose to use an actual image segmentation algorithm because this ensured grouping of spatially contiguous voxels into the same segment. This restricts neighboring voxels to move together and prevents unrealistic or incoherent jump moves from being proposed.

We observed that arbitrary tiling leads to consistently low PAR and extremely slow convergence. We believe this is because arbitrary tiling does not consider natural segments presented in the initial image and group voxels from spatially noncontiguous regions. On the other hand, more expensive and accurate segmentation routines did not significantly improve convergence rate or final image compared to (48). We also observed that a moderate amount of segments (around 100) are necessary to generate high quality final images, but more segments are discouraged as it required disproportionally higher processing time and it did not improve the final image.

Master Dictionary Construction.—We crucially assume that linear combinations of high frequency feature extracted from the initial solution are sufficient to generate a dictionary of plausible move proposals, because it is the high frequency component (edges, corners, noise) where the edge-preserving priors differ from conventional smoothness priors. Therefore, a set of moves encapsulating this high frequency content would be effective in removing noise and preserving edge-related details.

Therefore, we construct the master dictionary of high frequency information by applying a filter bank $F(\boldsymbol{x}, \theta)$ to the initial image. We extract high frequency information by subtracting the initial image from the filtered images. Let $\Omega = \{\theta_1, \theta_2, \ldots, \theta_n\}$ denote the set of parameters which could represent various frequency bands, orientations and textures (49). Then each of the n high frequency dictionary vector is simply computed as $\boldsymbol{z}^i = F(\boldsymbol{x}^0, \theta_i) - \boldsymbol{x}^0$. We employ bilateral filter bank (50) of varying parameters. We perform three-dimensional bilateral filtering separately on the real and imaginary components of $\boldsymbol{x}^0$ with parameters θ_i, and add up the separately filtered components to obtain the complex filtered

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

image $F(\boldsymbol{x}^0, \theta_i)$. Once $\boldsymbol{z}^i$ are extracted, we further impose a mutually disjoint tiling of these feature images based on segmentation S. Formally, given the set $Z = \{\boldsymbol{z}^i\}$ and set of image segments $S = \{\boldsymbol{s}^j\}$, we form the dictionary $V_0 = \{\boldsymbol{z}^i \cdot \boldsymbol{s}^j\}$ (element-wise multiplication). For notational simplicity we also denote by V_0 the matrix whose columns are given by the members of the set V_0.

Heuristic for generating move proposals.—As has been demonstrated in (51), plausible move proposals can significantly speed up convergence even if they cannot solve the non-convex objective directly. Step 5a generates a solution to Eq.6, which provides us the coefficients to linearly combine all dictionary vectors to generate plausible move proposals in step 5b. Our motivation is that Eq.6 can be considered a rough convex approximation to Eq.7 and is quickly solved with LSQR (52). In practice, move proposals resulting from Eq.6 yields high PAR per iteration. This is just a heuristic; individual dictionary elements could also serve as proposal moves, but at the cost of slower convergence. Note that the solution to Eq.6 is only used to generate plausible move proposals, not to solve the non-convex problem. For this limited purpose, we found that a simple L_2 function such as Eq.6 suffices. We stop the LSQR routine well before convergence, as absolute optimality of Eq.6 is unnecessary. Replacing the L_2 norm in Eq.6 by a L_1 norm produced slightly higher PAR since the L_1 norm is a closer convex approximation to Eq.7, but it required disproportionally higher processing time without commensurate image quality improvement, and was dropped.

Constructing unary and pairwise terms from Eq.7.—*S*olving Eq.7 with QPBO requires rewriting it as a summation of unary and pairwise terms. Denote $M^k = (V^k)^* \mathrm{E}^* EV^k$. The data fidelity term in Eq.7 is expressed as

$$\left\| E\left(\boldsymbol{x}^{k-1} + V^k \boldsymbol{b}\right) - \boldsymbol{y} \right\|^2 = \boldsymbol{b}^T \left(V^k\right)^* E^* EV^k \boldsymbol{b} + 2\left(E\boldsymbol{x}^{k-1} - \boldsymbol{y}\right)^* EV^k \boldsymbol{b} + \left(E\boldsymbol{x}^{k-1} - \boldsymbol{y}\right)^* \left(E\boldsymbol{x}^{k-1} - \boldsymbol{y}\right)$$
$$= \sum_i b_i \left(M^k_{i,\,i} + c_i\right) + \sum_{i,\,j}^{i \neq j} b_i M^k_{i,\,j} b_j + constant$$

where c_i is the ith element of the vector $2(E\boldsymbol{x}^{k-1} - \boldsymbol{y})^* EV^k$ and $constant = (E\boldsymbol{x}^{k-1} - \boldsymbol{y})^* (E\boldsymbol{x}^{k-1} - \boldsymbol{y})$. The first summation contains all unary terms and the second summation contains all pairwise terms. Note that $\sum_i b_i M^k_{i,i} b_i = \sum_i b_i M^k_{i,i}$ as b_i is binary.

M^k is efficiently computed from the precomputed M^0. Recall that we produce n dictionary vectors for each of the J segments. Let $\boldsymbol{g}^j$ be a Jn-by-1 binary vector such that $g^j_i = \begin{cases} 1, & (j-1)n < i \leq jn \\ 0, & otherwise \end{cases}$, for $1 \leq j \leq J$. Let $\tilde{A} = [\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_J]$ where $\tilde{\boldsymbol{\alpha}}_j = \boldsymbol{g}^j \cdot \tilde{\boldsymbol{\alpha}}$. It is not difficult to verify that $V^k = V_0 \tilde{A}$. Therefore, $M^k = (V^k)^* E^* EV^k = \tilde{A}^* (V_0^* E^* EV_0) \tilde{A} = \tilde{A}^* M^0 \tilde{A}$, which is fast to compute as the dimensions of these matrices are small and $\tilde{A}$ is sparse and structured. As can be clearly seen, the reason that we use predefined dictionaries and segmentation instead of using more dynamic elements such as iteratively changing dictionaries, segments, or the technique proposed in (51) is because introducing these dynamic elements require a computation of $(V^k)^* E^* EV^k$ from scratch per iteration, which is prohibitively costly.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As columns of V^k are mutually disjoint, let $\sigma(i)$ denote the column index of the only non-zero entry in the i-th row of V^k. The smoothness term is expressed as

$$G_{EP}\left(\left(\boldsymbol{x}^{k-1}+V^k\boldsymbol{b}\right)\right)=\lambda\sum_{i,\ j}\min\left(\left|\left(x_i+\sum\nolimits_m V^k_{i,\ m}b_m\right)-\left(x_j+\sum\nolimits_n V^k_{j,\ n}b_n\right)\right|^p,\ T\right)$$
$$=\lambda\sum\nolimits_{i,\ j}\min\left(\left|(x_i+V^k_{i,\ \sigma(i)}b_{\sigma(i)})-(x_j+V^k_{j,\ \sigma(j)}b_{\sigma(j)})\right|^p,\ T\right)$$
$$=\lambda\sum\nolimits^{\sigma(i)=\sigma(j)}_{i,\ j}\min\left(\left|(x_i-x_j)+\left(V^k_{i,\ \sigma(i)}-V^k_{j,\ \sigma(i)}\right)b_{\sigma(i)}\right|^p,\ T\right)$$
$$+\ \lambda\sum\nolimits^{\sigma(i)\neq\sigma(j)}_{i,\ j}\min\left(\left|(x_i-x_j)+\left(V^k_{i,\ \sigma(i)}b_{\sigma(i)}-V^k_{j,\ \sigma(j)}b_{\sigma(j)}\right)\right|^p,\ T\right)$$

The first summation contains all unary terms and the second summation contains all pairwise terms. We thus express Eq.7 as a summation of constant, unary and pairwise terms, and QPBO can be applied to solve the resulting problem.

Source Code.—A Matlab implementation of the above algorithm is available at https://github.com/jiexunxu/Fast-MRI-Imaging-via-Graphcut

METHODS

Structural brain MRI from four healthy volunteers was acquired at 3T on Siemens Skyra whole-body scanner (Siemens, Erlangen, Germany) using a standard 16 channels head matrix coil. T1-weighted images were acquired using a three-dimensional Magnetization-Prepared RApid Gradient Echo (MPRAGE) sequence (TE = 3ms, TI = 1000ms, TR = 2000ms, flip angle = 9°, 1 mm^3 isotropic voxel, acquisition time = 8min34 sec). The data dimensions were 256-by-176-by-256 voxels. Coronary MR Angiography using a 1.5T GE Excite 12 MR scanner was performed on a healthy volunteer. A commercially available eight-element phased array cardiac coil (4 anterior and 4 posterior elements) was used for signal reception. The subject was imaged supine with vector electrocardiography gating for cardiac synchronization. An ECG-triggered segmented k-space navigator gated 3D steady state free precession (SSFP) 3D pulse sequence was used at full sampling density. The imaging parameters were as follows: TR = 4.0 ms, TE = 1.2 ms, flip angle = 60°, readout bandwidth = 62.5 kHz, slice thickness = 3 mm, 16 slices, FOV = 26 cm, in-plane resolution = 1.0mm × 1.0mm, 32 partial echoes per heartbeat (corresponding to an acquisition window of 128ms), and centric view ordering along k_z. To allow quantitative and qualitative performance evaluations, we acquired all data at full resolution with no acceleration; these served as reference images.

The aliased images were simulated by random k-space under-sampling in the axial plane. The k-space views perpendicular to the axial plane were fully sampled. The under-sampling strategy employed was variable density Poisson sampling used in (16). We used the self-calibrating strategy for sensitivity estimation, whereby a small, central, circular shaped area (2.2% of the area of each slice) was acquired at full density to estimate low-frequency (relative) sensitivity maps. These scans were zero-padded to full resolution and inverse Fourier transformed to the image domain. In the image domain, these sensitivities were dilated with a disk structure element to obtain better sensitivity maps.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Algorithmic parameters were chosen empirically. The ratio between the data term energy and smoothness term energy was 1:0.6. The truncation term was 0.3 * $\|\mathbf{x}^0\|_\infty$. The exponent p in $G_{EP}((\mathbf{x}^0 + V_0\boldsymbol{\alpha}))$ was 0.8. We constructed 800 vectors in the master dictionary with around 100 segments. Most parameters are fixed through all experiments. A few parameters, such as the weighting between data-fidelity term and EPP prior, changed for different acceleration factors. We also observed that our experimental result was not sensitive to a reasonably small perturbation to these parameters, suggesting that our algorithm was not sensitive to parameter changes. Regularized SENSE and CS were implemented for comparison. The regularizer in regularized SENSE was simply the L_2 norm on the image itself. We experimented with two different CS algorithms. We replaced $G_{EP}(\boldsymbol{x})$ in Eq.3 with an l1-norm, and solved the resulting objective with SPGL1 version 2.0 (53). We also used l1-SPIRiT (54) as another popular CS algorithm in recent works such as (55) and (56) In all three algorithms, parameters are tuned manually and the best results we got were presented.

RESULTS

Convergence Results

We ran our algorithm at 5x simulated acceleration on four brain datasets. Fig 2 plots the level of energy of Eq.5 achieved at various iterations. We tested all four algorithms on the Dell PowerEdge R910 system. All algorithms were implemented in Matlab. Our proposed algorithm took 1.5~2 hours to finish from step (1) to (6), whereas CS method took around 3 hours.

In step 4 of our algorithm, we precomputed a dense matrix $M^0 = V_0^* E^* E V_0$ before our graph-cut based minimization. While computing a dense matrix is discouraged in general, in our method, we have full control over the dimension of this matrix. Its dimension is the number of vectors in the master dictionary. Precomputing this matrix takes around 15 minutes, which is not a major bottleneck compared to the overall runtime. Furthermore, we expect that the most attractive use case of our method will be in serial imaging during the same session, e.g. multiple modalities or multiple time points. This may also apply to multiple subjects as long as they are using the exact same coils/set up (although there are practical challenges in that). In these cases, the pre-compute time will be amortized over many acquisitions.

Quantitative Results

One of the main applications of 3D T1 brain scans is for assessing regional atrophy patterns; hence, any reconstruction algorithm should preserve sufficient tissue contrast to allow error-free segmentation, registration and voluming. We performed downstream volumetric analysis on 6x reconstructed brain volumes with SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), for all four algorithms: SENSE, SPGL1, l1-SPIRiT and Graphcut. Briefly, we parcellated the grey-matter regions of the reconstructed and reference images into 116 regions obtained from the Anatomic Annotated Labeled (AAL) atlas. For each region, if the reference image reported a volume of V_0 and an under-sampled image reported a volume of V, then we computed the Relative Volumetric Difference (RVD) as $|V - V_0|/V_0$. Table 1 reports the average of these 116 RVD values.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We performed SENSE, CS, and Graphcut reconstructions at simulated acceleration factors from 3x to 13x on four brain datasets. For each reconstruction, we calculated the normalized mean-square errors (NMSE) against the corresponding reference image. A plot of the NMSE against acceleration factors is shown in Fig 3.

We also demonstrated the stability of our algorithm by artificially adding uncorrelated Gaussian noise to coil outputs at different noise levels. We calculated the NMSE of the reconstructed images at 5x acceleration. A plot of the NMSE against noise level is shown in Fig 4.

Qualitative Results

Fig 5~Fig 7 show simulated results on three datasets, reconstructed using SENSE, SPGL1, l1-SPIRiT and Graphcut. All parameters were fixed through these three experiments. The bottom row show the corresponding difference images with enhanced contrast. The NMSE for each difference image is also displayed. Fig 5 and Fig 6 simulate 4x acceleration, and Fig 7 simulates 5x acceleration.

To investigate how far we can push the acceleration factor, we attempted a reconstruction at 7x acceleration (shown in Fig 8). We used CS image as our initial image because SENSE image was too noisy to work with at 7x.

In Fig 9 we reported preliminary experiments on the MR coronary Angiography dataset at 7x acceleration. Although this example, being a multi-slice acquisition, is not directly amenable to random under-sampling, our results should serve as a motivation for future applications in torso imaging.

DISCUSSION AND CONCLUSION

Discussion of Results

In this paper we have shown that the use of combinatorial optimization methods relying on graphcuts can successfully solve a challenging class of image reconstruction problems. The ability of the algorithm to handle non-convex regularization terms imparts it certain desirable properties from a practical point of view. In particular, the reconstructed images show better edge preservation and higher SNR compared to many other methods. These benefits come at the cost of higher complexity of the algorithm. In particular, the ability to handle EPPs as well as non-random acquisitions requires the constriction of a dictionary of image features. The algorithm then performs moves along these feature vectors, solving a succession of binary optimization problems.

In Fig 2, the energy is reduced drastically in the first few iterations, and decreases slowly thereafter – a typical behavior of graphcut algorithms. Such fast convergence allows our algorithm to beat CS in runtime. In the quantitative experiments, our algorithm (Graphcut) reports the lowest or close to the lowest average RVD and best overall NMSE scores compared to the other three methods. In Fig 5~Fig 7, results from our algorithm suppress more noise compared to competing methods, as can be seen visually and can be demonstrated by the difference images and NMSE values. Our results also produce slightly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sharper edges. Due to the EPP prior, our results demonstrate piecewise smoothness: noises are sufficiently suppressed in both the grey matter and white matter while edge sharpness is preserved. Although it is unclear whether highly accelerated images presented in Fig 8 can be diagnostically useful, the improved visual quality of our result over competing methods indicates our algorithm's potential and promise. Fig 9, an application of our algorithm on another body part, also demonstrates improved noise suppression. This experiment shows that although our algorithm was specifically customized for neuroimaging, it can be profitably applied to other targets.

Limitations and Future Steps

One limitation of our algorithm is that the results have some staircasing artifacts, such as the white blocks in Fig 8. This is a well-known feature of the EPP prior. Such artifacts can also been seen in our previous EPIGRAM work. Staircasing represents one of the compromises made to achieve sharp edge boundaries. It is important to know this trade-off, and the user should carefully choose whether and how much they wish to enforce piece-wise smoothness, versus normal smoothness. While we believe that the staircasing effect is unlikely to lead to misdiagnosis of lesions, we must admit this as a possibility.

Our algorithm performance also depends on the quality of the initial image, and is prone to propagate artifacts. Another limitation is the need to generate good dictionary elements from a noisy and aliased initial image. The dictionary itself does not have to be perfect; however, the effect of various features were not thoroughly investigated in the current work. In the future, we will extensively explore other dictionary frames and features, such as wavelet and anisotropic diffusion. We consider such exploration to be an orthogonal problem to the one we are trying to solve in this paper. We will also implement atlas-based feature detection, but co-registering the image to anatomic atlases, which may provide higher quality edge or texture features than the initial image by itself, and may help against the staircasing artifact present in our results.

Another potential for improvement lies in automatic parameter tuning. Currently we empirically pick the best parameter set for each imaging sequence, but thence onward no further tuning is required for individual subjects. Although our algorithm is not sensitive to parameter changes, they still need to be optimized by hand for different MR modalities and scanners. Therefore, automatic or adaptive parameter tuning would be beneficial for wider applicability.

We may also try to minimize the objective function with our EPP prior with other methods. Combinatorial algorithms like ours are not the only way to address non-convexity. Even continuous optimization methods can give reasonable solutions of non-convex problems, and can sometimes give a favorable trade-off between complexity, execution speed and performance. For example, a "greedy" approach, as detailed in Equation 3 in (57), can be used to minimize our EPP prior. Such method has been used in other publications for solving related non-convex problems. There are also published papers that solve $L_p\,(p < 1)$ norms with L_1 majorizers, such as (58), that we can use in our method. Additionally, some other published work, such as (59) and (60), solves the image reconstruction problem with different objective functions that are also non-convex. Depending on the specific formulation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

desired by the user, these methods might prove simpler or superior to ours. A comparison of our algorithm against these methods would be another future work.

Furthermore, this paper is mainly concerned with an efficient combinatorial formulation of the CS problem, rather than with image modeling; but there is room for more complicated, higher-order priors involving non-local or patch-based features, which could complement the conservative, low-level neighborhood priors used here. Our dictionary-based approach is well placed to incorporate such models. However, higher-order or overly strong reference-based priors meet with stronger resistance in the medical imaging field, and rightly so, due to the possibility of hallucinating artifacts. In future work we will consider whether some of these features could safely be deployed in MR.

Conclusion

We have presented a novel algorithm for combining CS and PI, and introduced a highly non-convex EPP prior to show that the minimizing the corresponding objective function can result in a higher quality MRI image. We also presented an algorithm, based on graphcut-based minimization techniques, to overcome the non-convexity of the problem. This provides an entirely new avenue for handling the challenges arising from CS. Although much more future work will be required to thoroughly quantify and characterize both our EPP prior and our algorithm on a variety of imaging applications, we believe our presented results provide preliminary but encouraging assessment of the potential of this novel approach. Our use of graph-cut based minimization techniques can also be a potentially new and useful algorithmic addition to the community.

ACKNOLEDGEMENTS

This work was partially supported by NIH grants R01 NS075425, P41 EB015904 and R21EB008138, R01 NS092802 and RF1 AG062196. The authors would like to thank David Bindel, Ramin Zabih, Farras Abdelnour and Pascal Spincemaille for their helpful discussions.

List of Abbreviations:

PI	Parallel Imaging
CS	Compressive Sensing
SENSE	Sensitivity Encoding
EPP	Edge Preserving Priors
EPIGRAM	Edge-preserving Parallel Imaging with Graph-cut Minimization
QPBO	Quadratic Pseudo-Boolean Optimization
PAR	Proposal Acceptance Ratio

REFERENCES

(1). Pruessmann KP, Weiger M, Scheidegger MB, and Boesiger P, “SENSE: sensitivity encoding for fast MRI.,” Magnetic resonance in medicine : official journal of the Society of Magnetic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 42, no. 5, pp. 952–62, Nov. 1999.

(2). Weiger M, Pruessmann KP, and Boesiger P, “2D SENSE for faster 3D MRI.,” Magma (New York, N.Y.), vol. 14, no. 1, pp. 10–9, Mar. 2002.

(3). Pruessmann KP, Weiger M, Börnert P, and Boesiger P, “Advances in sensitivity encoding with arbitrary k-space trajectories.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 46, no. 4, pp. 638–51, Oct. 2001.

(4). Sodickson DK and Manning WJ, “Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays,” Magnetic Resonance in Medicine, vol. 38, no. 4, pp. 591–603, Oct. 1997. [PubMed: 9324327]

(5). Sodickson DK, McKenzie CA, Ohliger MA, Yeh EN, and Price MD, “Recent advances in image reconstruction, coil sensitivity calibration, and coil array design for SMASH and generalized parallel MRI.,” Magma (New York, N.Y.), vol. 13, no. 3, pp. 158–63, Jan. 2002.

(6). Griswold MA et al. , “Generalized autocalibrating partially parallel acquisitions (GRAPPA).,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 47, no. 6, pp. 1202–10, Jun. 2002.

(7). Wang Y, “Description of parallel imaging in MRI using multiple coils.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 44, no. 3, pp. 495–9, Sep. 2000.

(8). Lin F-H, Kwong KK, Belliveau JW, and Wald LL, “Parallel imaging reconstruction using automatic regularization.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 51, no. 3, pp. 559–67, Mar. 2004.

(9). Bammer R et al. , “Diffusion tensor imaging using single-shot SENSE-EPI.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 48, no. 1, pp. 128–36, Jul. 2002.

(10). Haldar JP, Hernando D, Song S-K, and Liang Z-P, “Anatomically constrained reconstruction from noisy data.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 59, no. 4, pp. 810–8, Apr. 2008.

(11). Lustig M, Lustig M, Donoho DL, and Pauly JM, “Rapid MR imaging with Compressed Sensing and randomly under-sampled 3DFT trajectories,” IN PROC. 14TH, 2006.

(12). Lustig M, Donoho DL, Santos JM, and Pauly JM, “Compressed Sensing MRI,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72–82, Mar. 2008.

(13). Lustig M, Donoho D, and Pauly JM, “Sparse MRI: The application of compressed sensing for rapid MR imaging.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–95, Dec. 2007.

(14). Boykov Y, Veksler O, and Zabih R, “Fast approximate energy minimization via graph cuts,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222–1239, 2001.

(15). Otazo R, Kim D, Axel L, and Sodickson DK, “Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 64, no. 3, pp. 767–76, Sep. 2010.

(16). Vasanawala SS,Murphy MJ,Alley MT,Lai P,Keutzer K,John M Pauly Michael Lustig, “Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients.,” 2011, pp. 1039 – 1043.

(17). Tibshirani Robert, “Regression shrinkage and selection via the Lasso,” JSTOR: Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. No 1, pp. 267–288, 1996.

(18). Donoho DL, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(19). Tropp JA, “Just relax: Convex programming methods for subset selection and sparse approximation.”

(20). Poggio T, Torre V, and Koch C, “Computational vision and regularization theory,” Nature, vol. 317, no. 6035, pp. 314–319, Sep. 1985. [PubMed: 2413361]

(21). Li S, Markov Random Field Modeling in Computer Vision. Berlin: Springer-Verlag, 1995.

(22). Geman S and Geman D, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” pp. 452–472, Jun. 1990.

(23). Besag J, On the statistical Analysis of Dirty Pictures. J R Stat Soc Ser B, 1986, pp. 48:259–302.

(24). Chellappa R JA, Markov random fields: theory and applications. New York: Academic Press, 1993.

(25). Black MJ and Rangarajan A, “On the unification of line processes, outlier rejection, and robust statistics with applications in early vision,” International Journal of Computer Vision, vol. 19, no. 1, pp. 57–91, Jul. 1996.

(26). Black MJ and Anandan P, “A framework for the robust estimation of optical flow,” in 1993 (4th) International Conference on Computer Vision, pp. 231–236.

(27). Chartrand R, “Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data,” INT. SYMP. BIOMEDICAL IMAING, 2009.

(28). Raj A et al. , “Bayesian parallel imaging with edge-preserving priors.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 57, no. 1, pp. 8–21, Jan. 2007.

(29). Singh G et al. , “A fast Edge-preserving Bayesian reconstruction method for Parallel Imaging applications in cardiac MRI.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 65, no. 1, pp. 184–9, Jan. 2011.

(30). Yin P and Xin J, “Iterative L_1 minimization for non-convex compressed sensing,” Journal of Computational Mathematics, 35.10.4208, April. 2016.

(31). Ince T and Ögücü G, “Array failure diagnosis using nonconvex Compressed Sensing,” IEEE Antennas and Wireless Propagation Letters, vol. 15, pp. 992–995, Oct. 2015.

(32). Trzasko J and Manduca A, “Highly undersampled magnetic resonance image reconstruction via homotopic l(0) -minimization.,” IEEE transactions on medical imaging, vol. 28, no. 1, pp. 106–21, Jan. 2009. [PubMed: 19116193]

(33). Trzasko JD et al. , “Sparse-CAPR: highly accelerated 4D CE-MRA with parallel imaging and nonconvex compressive sensing.,” Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 66, no. 4, pp. 1019–32, Oct. 2011.

(34). Ajanthan T et al. , “Iteratively reweighted graph cut for multi-label MRFs with non-convex priors.,” The IEEE Conference on Computer Vision and Pattern Recognition, pp. 5144–5152, Jun. 2015.

(35). Rother C, Kolmogorov V, Lempitsky V, and Szummer M, “Optimizing binary MRFs via extended roof duality,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

(36). Bleyer M and Gelautz M, “Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions,” Signal Processing: Image Communication, vol. 22, no. 2, pp. 127–143, Feb. 2007.

(37). Lu F et al. ,, “Automatic 3D liver location and segmentation via convolutional neural network and graph cut,” International Journal of Computer Assisted Radiology and Surgery, vol. 12, no. 2, pp. 171–182, Sep. 2016. [PubMed: 27604760]

(38). Oesau S et al. , “Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut,” in ISPRS Journal of Photogrammetry and Remote Sensing, vol. 90, pp. 68–82, Apr. 2014

(39). Kolmogorov V and Zabih R, “What energy functions can be minimized via graph cuts?,” IEEE transactions on pattern analysis and machine intelligence, vol. 26, no. 2, pp. 147–59, Feb. 2004. [PubMed: 15376891]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(40). Felzenszwalb PF and Zabih R, “Dynamic programming and graph algorithms in computer vision.,” IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 4, pp. 721–40, Apr. 2011. [PubMed: 20660950]

(41). Veksler O, “Efficient graph-based energy minimization methods in Computer Vision.”

(42). Bi C, Wang H and Bao R, “SAR image change detection using regularized dictionary learning and fuzzy clustering,” in IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, 2014.

(43). Shang J, Huang Z, Li Q and Zhang T, “Sparse representations via learned dictionaries for x-ray angiogram image denoising”, Proc. SPIE 10610, MIPPR 2017: Parallel Processing of Images and Optimization Techniques; and Medical Imaging, 106100C, Mar. 2018

(44). Victor Lempitsky CR, “Fusion moves for Markov random field optimization.”

(45). Kleinberg J and Tardos É, Algorithm Design. Addison Wesley, 2005, p. 864.

(46). Kolmogorov V and Zabih R, “Computing visual correspondence with occlusions using graph cuts,” in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 2001, vol. 2, pp. 508–515.

(47). Kolmogorov V and Zabih R, “Multi-camera scene reconstruction via graph cuts,” pp. 82–96, May 2002.

(48). Felzenszwalb PF, Felzenszwalb PF, and Huttenlocher DP, “Efficient graph-based image segmentation,” INTERNATIONAL JOURNAL OF COMPUTER VISION, vol. 59, 2004.

(49). Vetterli M and Herley C, “Wavelets and filter banks: theory and design,” IEEE Transactions on Signal Processing, vol. 40, no. 9, pp. 2207–2232, 1992.

(50). Tomasi C and Manduchi R, “Bilateral filtering for gray and color images,” in Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), pp. 839–846.

(51). Ishikawa H, “Higher-order gradient descent by fusion-move graph cut,” in 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 568–574.

(52). Paige CC and Saunders MA, “LSQR: an algorithm for sparse linear equations and sparse least squares,” ACM Transactions on Mathematical Software, vol. 8, no. 1, pp. 43–71, Mar. 1982

(53). van den Berg E and Friedlander MP, “Probing the Pareto frontier for basis pursuit solutions.” 2008.

(54). Lustig M and Pauly JM, “SPIRiT: Iterative self-consistentparallel imaging reconstruction from arbitrary k-space,” Magnetic Resonance in Medicine, 451–71, 2010

(55). Hosseini SAH et al. , “Accelerated coronary MRI using 3D Spirit-Raki with sparsity regularization”, IEEE 16th International Symposium on Biomedical Imaging, 2019.

(56). Shimron E, Webb AG and Azhari H, “CORE-PI: Non-iterative convolution-based reconstruction for parallel MRI in the wavelet domain”, Medical Physics, 2019

(57). Ravishankar S, Nadakuditi R and Fessler J, “Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems”, IEEE Transactions on Computational Imaging, volume 3, issue 4, 2017

(58). Figueiredo M, Bioucas-Dias J and Nowak R, “Majorization-minimization algorithms for wavelet-based image restoration”, IEEE Transactions on Image Processing, volume 16, issue 12, 2007

(59). Ravishankar S and Bresler Y, “Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging”, SIAM Journal on Imaging Sciences, vol 8, issue 4, 2015

(60). Ravishankar S, Ye J and Fessler J, “Image reconstrution: from sparsity to data-adaptive methods and machine learning”, Proceedings of the IEEE, vol 109, issue 1, 2020

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

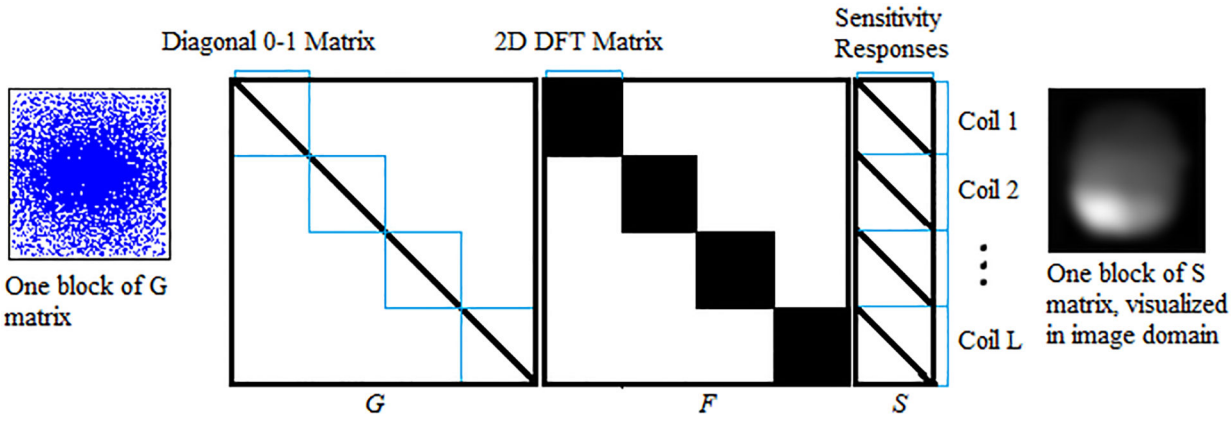

Fig 1:

Visual explanation of Eq. 1. The G matrix consists of identical blocks of 0–1 diagonal matrices. The diagonal in each block is a column vectorization of the weighted random sampling matrix visualized on the left. A blue entry is 1 and a white entry is 0. The S matrix is a block diagonal matrix. The diagonal in the i-th block is a column vectorization of the sensitivity response of the i-th coil.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

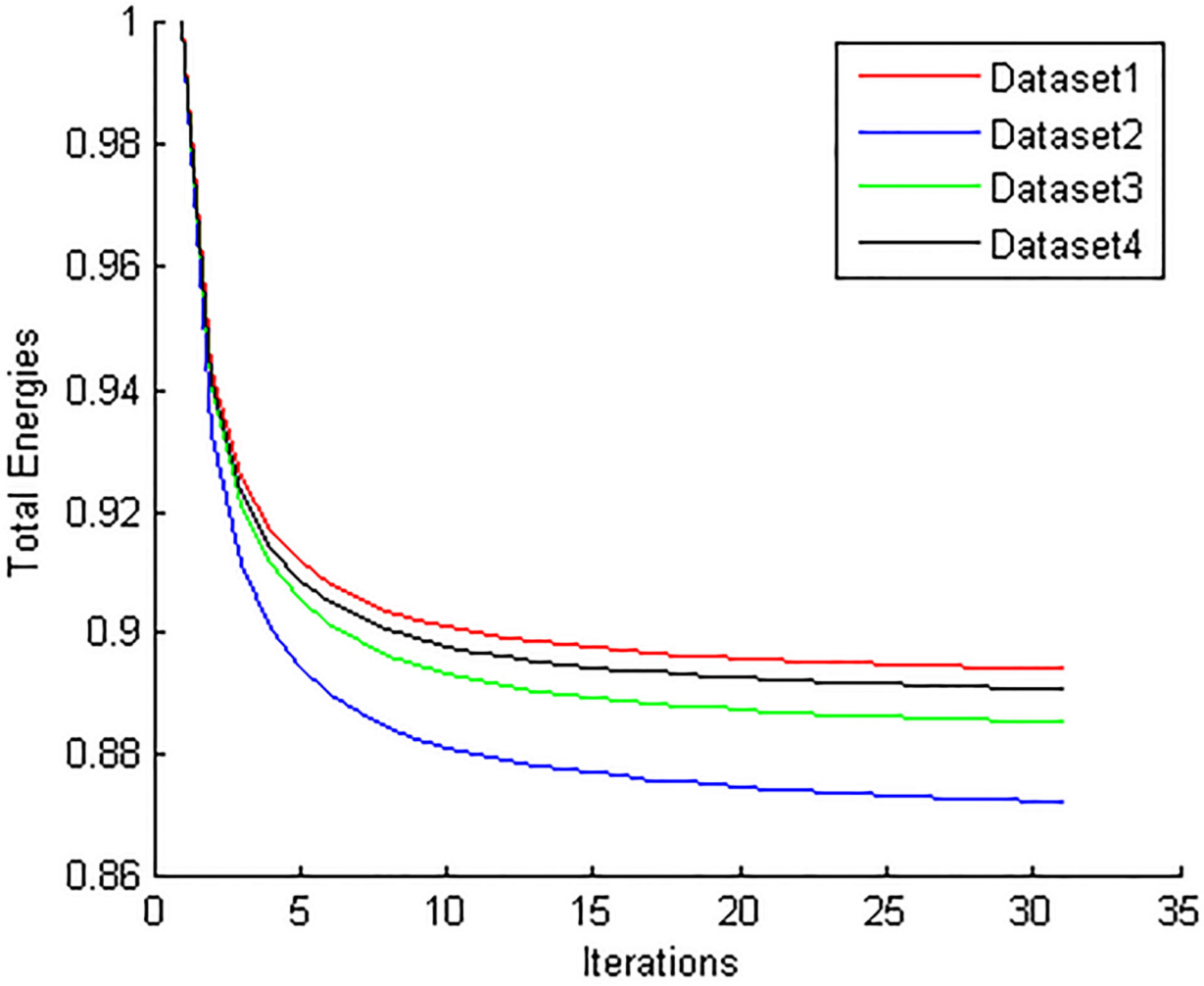

Fig 2:

Plot of total energy against graphcut iterations for four datasets. The energy decreases much faster within the first few iterations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

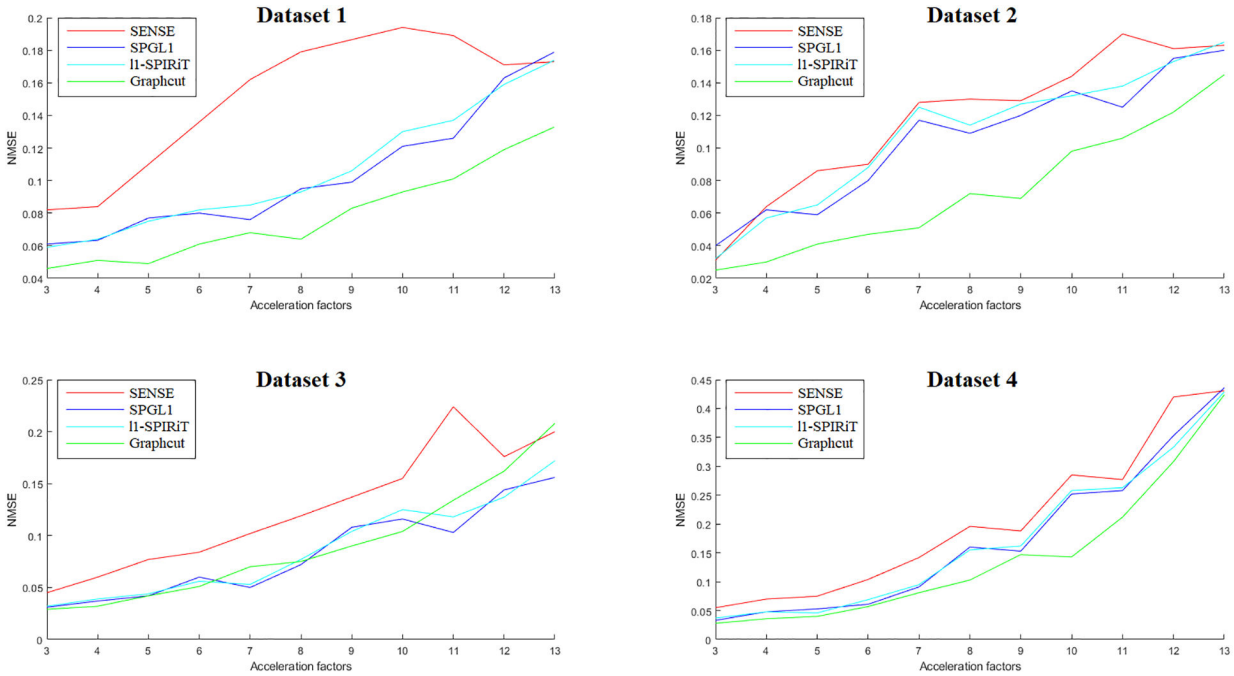

Fig 3:
Plot of NMSE against various acceleration factors

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

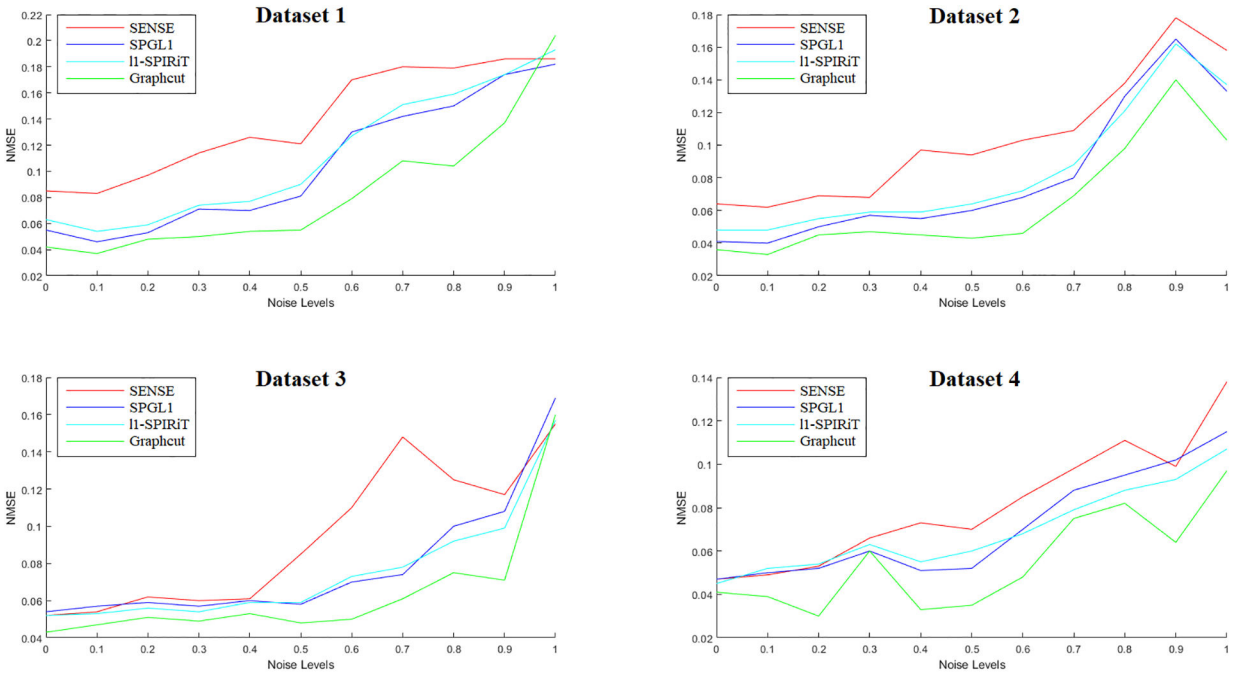

Fig 4:
Plot of NMSE against different noise levels in the coil data

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

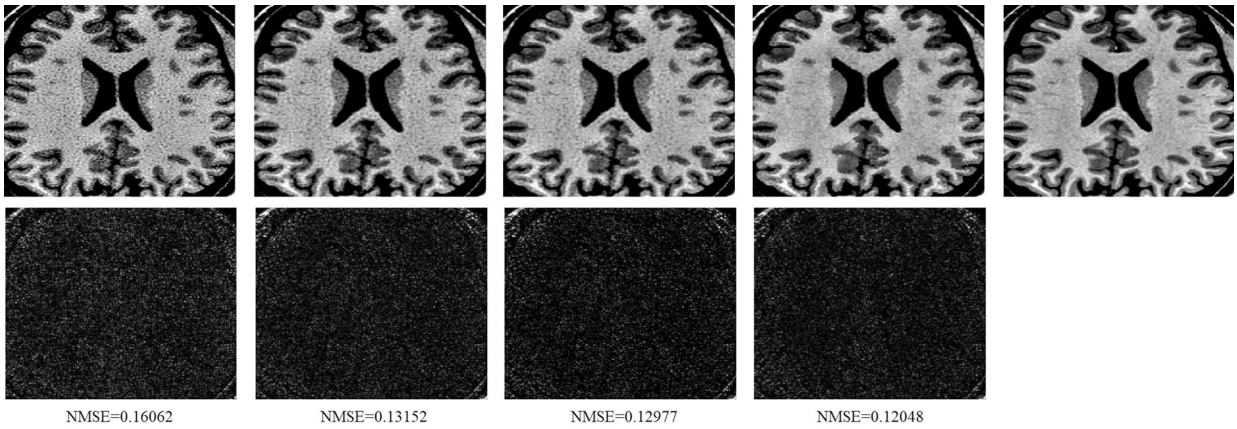

Fig 5:

SENSE (1st column), SPGL1 (2nd column), l1-SPIRiT (3rd column) and Graphcut (4th column) reconstruction and reference image at 4x acceleration. The bottom row contains difference image with properly enhanced brightness and contrast

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

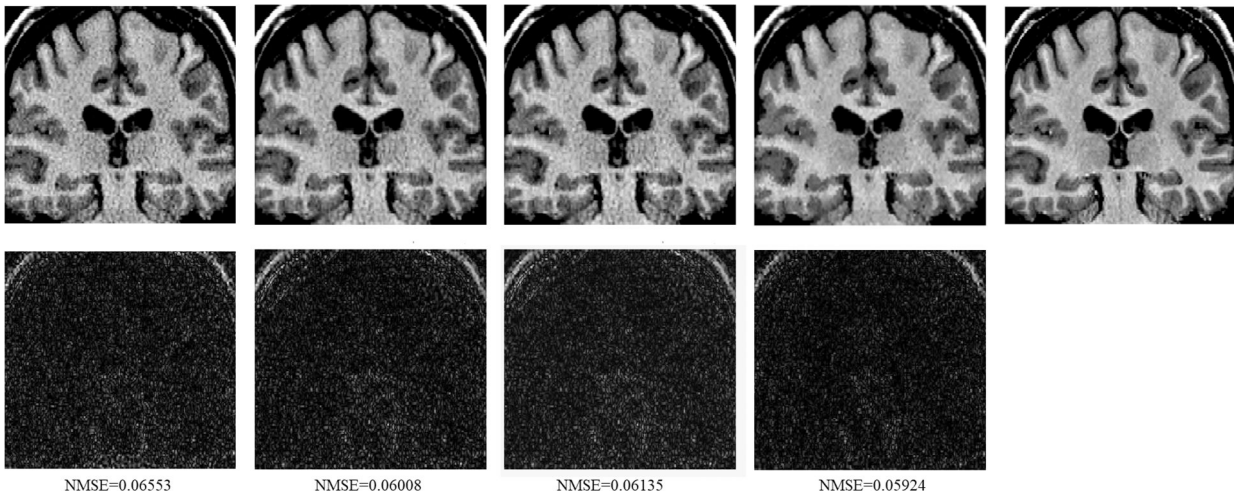

Fig 6:

SENSE (1st column), SPGL1 (2nd column), l1-SPIRiT (3rd column) and Graphcut (4th column) reconstruction and reference image at 4x acceleration. The bottom row contains difference image with properly enhanced brightness and contrast

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

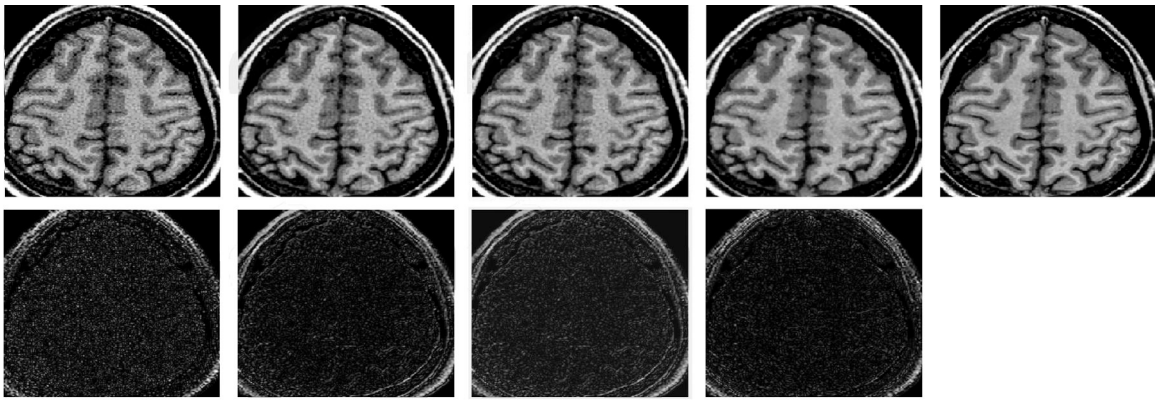

Fig 7:

SENSE (1st column), SPGL1 (2nd column), l1-SPIRiT (3rd column) and Graphcut (4th column) reconstruction and reference image at 5x acceleration. The bottom row contains difference image with properly enhanced brightness and contrast

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

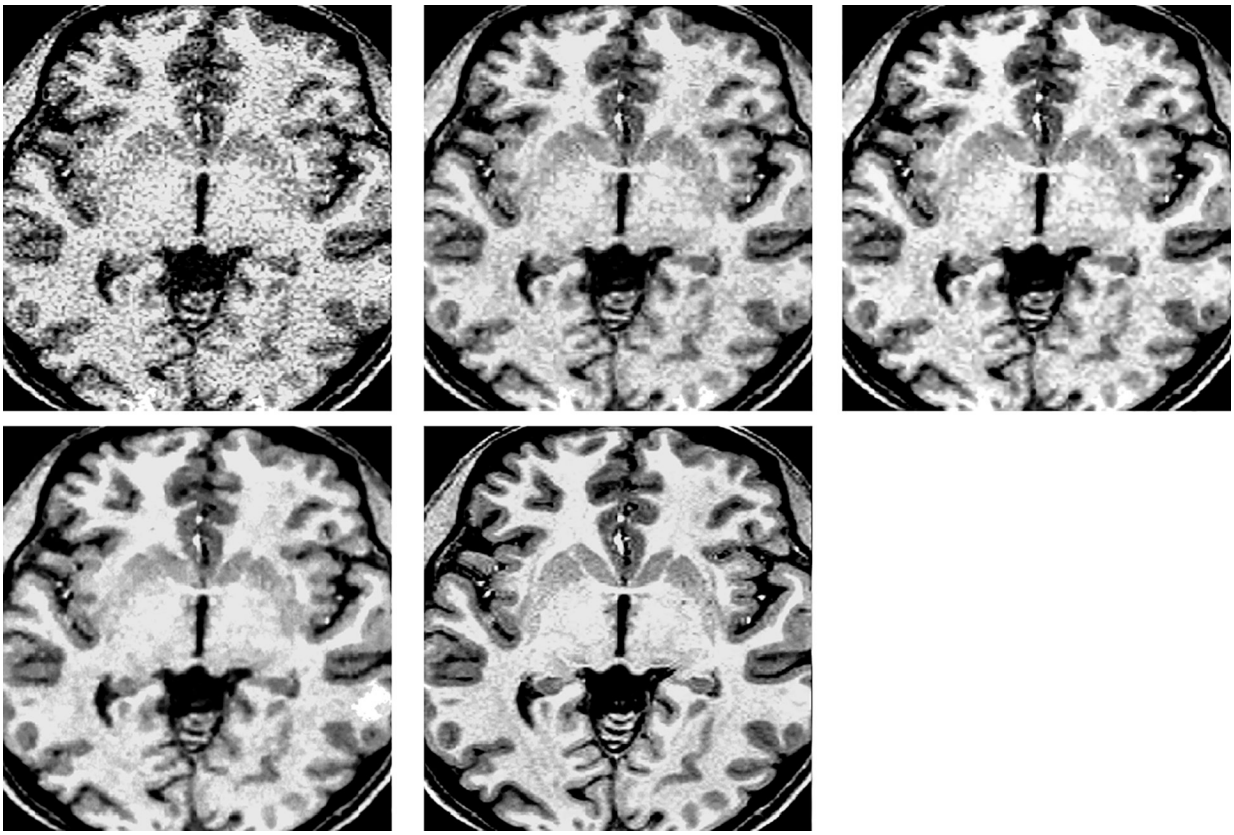

Fig 8:
SENSE (top left), SPGL1 (top middle), l1-SPIRiT (top right), Graphcut (bottom left) reconstruction and reference image (bottom right) at 7x acceleration.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

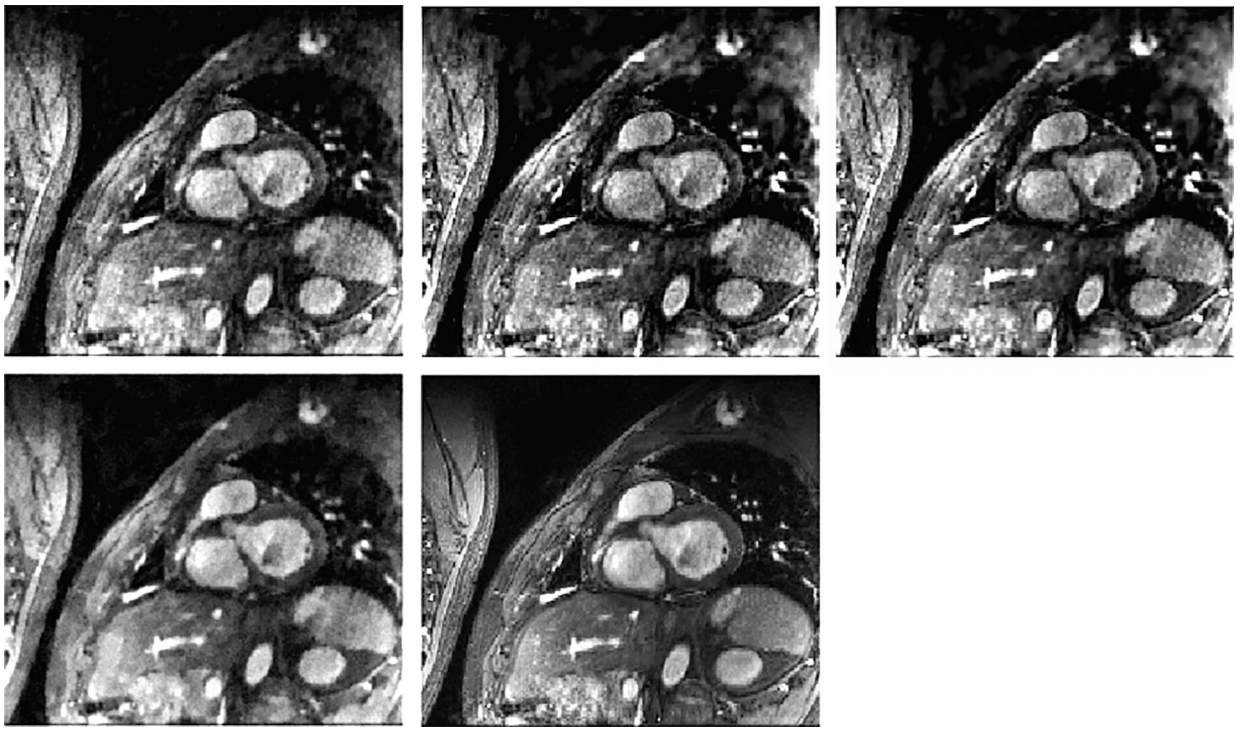

Fig 9:
SENSE (top left), SPGL1 (top middle), l1-SPIRiT (top right and Graphcut (bottom left) reconstruction and reference image (bottom right) at 7x acceleration.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

SPM brain volumetric analysis: Average Relative Volume Difference scores for SENSE, CS and Graphcut on four brain MPRAGE scans. The lowest error is shown in **boldface**.

	SENSE	SPGL1	l1-SPIRiT	Graphcut
Dataset 1	0.0745	0.0612	0.0588	**0.0458**
Dataset 2	0.0711	**0.0582**	0.0641	0.0630
Dataset 3	0.1220	0.0704	0.0753	**0.0536**
Dataset 4	0.0737	0.0673	0.0599	**0.0547**