Magnetic Resonance Materials in Physics, Biology and Medicine

Fast Reduction of Undersampling Artifacts in Radial MR Angiography with 3D Total Variation on Graphics Hardware

Florian Knoll[1]*, Markus Unger[2], Clemens Diwoky[1], Christian Clason[3], Thomas Pock[2] and Rudolf Stollberger[1]

[1]Institute of Medical Engineering, Graz University of Technology, Kronesgasse 5, A-8010 Graz, Austria

[2]Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, A-8010 Graz, Austria

[3]Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria

*Correspondence to: Florian Knoll,
Graz University of Technology,
Institute of Medical Engineering,
Kronesgasse 5, A-8010 Graz, Austria.
Phone: +43-316-873-5375 , Fax: +43-316-873-7890.
E-mail: florian.knoll@tugraz.at

Number of Words (Abstract) = 275
Number of Words (Body) = approximately 6000
Number of Figures: 6
Number of Tables: 2
Number of References: 32

Abstract *Objective:* Subsampling of radially encoded MRI acquisitions in combination with sparsity promoting methods opened a door to significantly increased imaging speed, which is crucial for many important clinical applications. In particular, it has been shown recently that total variation (TV) regularization efficiently reduces undersampling artifacts. The drawback of the method is the long reconstruction time which makes it impossible to use in daily clinical practice, especially if the TV optimization problem has to be solved repeatedly to select a proper regularization parameter.

Materials and Methods: The goal of this work was to show that for the case of MR-Angiography, TV filtering can be performed as a post-processing step, in contrast to the common approach of integrating TV penalties in the image reconstruction process. With this approach it is possible to use TV algorithms with data fidelity terms in image space, which can be implemented very efficiently on graphic processing units (GPUs). The combination of a special radial sampling trajectory and a full 3D formulation of the TV minimization problem is crucial for the effectiveness of the artifact elimination process.

Results and Conclusion: The computation times of GPU-TV show that interactive elimination of undersampling artifacts is possible even for large volume data sets, in particular allowing the interactive determination of the regularization parameter. Results from phantom measurements and in vivo angiography data sets show that 3D TV, together with the proposed sampling trajectory, leads to pronounced improvements in image quality. However, while artifact removal was very efficient for angiography data sets in this work, it cannot be expected that the proposed method of TV post-processing will work for arbitrary types of scans.

Keywords Angiography · Accelerated Imaging · Radial Sampling · Total Variation · GPU Computing

1 Introduction

The time window for data acquisition in contrast enhanced MR angiography (CE-MRA) is limited due to the passage of the contrast agent. Additionally, a high spatial resolution is needed to visualize small vessels, while high temporal resolution is necessary to capture the dynamics of the contrast agent bolus. However, due to fundamental properties of MRI, there is always a trade off between spatial and temporal resolution. As MRI data acquisition is sequential, it can be accelerated by reducing the number of measurement steps, but this leads to artifacts in the reconstructed images because the Nyquist criterion is violated. There are different strategies to mitigate these artifacts such as parallel imaging [25], [14] or methods that exploit spatio-temporal correlations [18], [29], [19].

During the last years, reconstruction strategies were formulated that use tailored acquisition strategies like radial sampling or randomized 3D sampling patterns and include a priori knowledge about the imaged objects during the reconstruction process [17], [3]. Examples for a priori knowledge are sparsity of the image, or total variation (TV) based methods which assume that the image consists of areas that are piecewise constant. Mathematically, these additional assumptions are introduced by reformulating image reconstruction as a constrained optimization problem. These algorithms allow the reconstruction of high quality images from highly undersampled data sets, but at the price of long computation time. While this is not a severe limitation during research, it makes them currently impossible to use in daily clinical practice.

Hansen et al. [15] and Sorensen et al. [28] have shown recently that it is possible to use the massively parallel streaming processor architecture of modern graphic processing units (GPUs) to speed up image reconstruction for parallel imaging and for the reconstruction of non-Cartesian data. The goal of this work was to show that for the special case of CE-MR angiography data sets, which feature a high contrast to noise ratio, it is possible to use TV algorithms with data fidelity terms in image space, instead of integrating TV regularization in the image reconstruction process. For these algorithms, primal-dual based TV formulations can be implemented very efficiently on the GPU, with computation times that make interactive elimination of undersampling artifacts possible. This is especially important for the interactive determination of the regularization parameter because a proper choice usually depends on patient-specific conditions like the geometry of the imaged anatomy, and therefore predefined settings may not deliver optimal results. It is also shown in this paper that by using a special type of radial sampling, 3D a priori information can be included in the regularization process, which significantly improves image quality. While this leads to an even higher computational complexity, it becomes feasible with the GPU implementation.

2 Theory

2.1 Undersampled Radial Imaging and Total Variation

It is well known [2] that in order to reconstruct an $n \times n$ image matrix from a fully sampled radial data set, $\frac{\pi}{2} \cdot n$ radial projections have to be acquired. Reducing the number of projections accelerates data acquisition, but leads to characteristic streaking artifacts in the reconstruction. It was shown by Block et al. [3] that these streaking artifacts can be reduced efficiently by using a TV regularization in conjunction with a data fidelity term in k-space, a strategy that is comparable with compressed sensing approaches [17]. In contrast, our method is based on the original TV approach formulated in image space. TV based min-

imization problems were originally designed for elimination of Gaussian noise, and were first used by Rudin, Osher and Fatemi [27] in 1992. The denoising model using a TV regularization together with a L^2 data term in image space is therefore often referred to as the ROF model. It is defined as the following minimization problem:

$$\min_u \left\{ \int_\Omega |\nabla u|\, dx + \frac{\lambda}{2} \int_\Omega (u - f)^2\, dx \right\}, \tag{1}$$

where f is the original, corrupted image data. In the original work [27], the corruption is due to Gaussian noise, while in our case, it comes from streaking artifacts due to subsampling. The minimizer u is the reconstructed image, and Ω is the image domain. The regularization parameter λ controls the balance between artifact removal – minimizing the TV penalty – and faithfulness to the original image – minimizing the data fidelity term, – and so the proper choice of this parameter is a major challenge in any regularization based reconstruction method. The L^2 norm of the data fidelity term makes the removal of structures contrast dependent: Low contrast regions are removed, but strong contrast regions of the same size are kept. While this is desirable for high-contrast data such as CE-MR angiography, it is not appropriate for other data sets with low contrast features. For such applications, it is possible to extend the proposed approach to include an L^1 data fidelity term, which is better suited due to its contrast invariance. The TV regularization has the advantage of removing noise while preserving sharp edges in the image. Therefore, vessels with their strong contrast-to-noise-ratio are preserved, while undersampling artifacts are efficiently removed. The main difference to the method in [3] is that we use a two step procedure: First, a Fourier transform is applied to the MRI raw data, giving a reconstructed image containing aliasing artifacts. This image is then filtered through TV minimization. In contrast, Block et al. use a single step procedure where the TV based regularization is integrated in the image reconstruction process.

The minimization of the ROF model is a well studied problem. As the functional defined in (1) is convex, the unique global minimizer can be calculated. In the original formulation, this was done using explicit time marching [27]. Other methods employ a linearization of the Euler-Lagrange equation [30], [8]. Duality based methods have shown greatly improved performance and were, among others, proposed in [9], [6], [5] and [7]. Recently, very fast primal-dual (PDU) approaches were proposed by Zhu et al. in [31] and [32], although they were already used for saddle point problems by Popov in [24]. In [13], a Split Bregman algorithm is used for the minimization. Discrete methods using graph cuts also can be used to solve the ROF model [12]. As shown in [23], continuous methods have the advantage of their inherent parallelization potential, a low memory consumption and no discretization artifacts.

Recently, continuous methods have become applicable for large 3D data sets by implementing them on the GPU [23]. It is shown in this paper that by adapting the PDU approach from [31] to 3D and by implementing it on the GPU, even large volume data sets can be processed in a reasonable time.

To benefit from the TV constraint in the third dimension, the streaking artifacts must appear in a different pattern in adjacent kz-planes. This can be achieved easily with a modification of the sampling trajectory. Currently, most radial acquisition patterns use a so called "stack of stars" trajectory [22] with radial sampling in the xy-plane and Cartesian encoding or multi slice acquisition in the z-direction. This sampling pattern uses the same projection angles in all kz-planes, which creates streaking artifacts in the image volume which are two-dimensional (specifically, independent of z) and therefore difficult to eliminate by
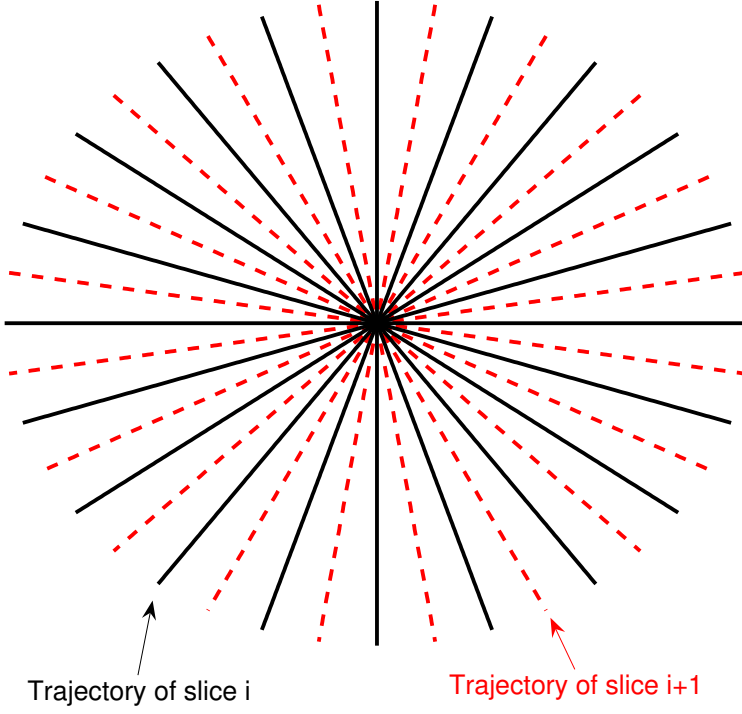
Trajectory of slice i   Trajectory of slice i+1

Fig. 1 Stack of stars trajectory with shifted projection angles, 10 radial projections are displayed.

TV minimization (which is chosen precisely for the fact that it preserves jumps over two-dimensional sets, such as artery boundaries). On the other hand, if the projection angles of adjacent kz-planes are shifted by $\frac{\pi}{2k}$, where k is the number of projections (see Figure 1), the aliasing artifacts have a different, line-like, structure which is much more easily removed. This should be compared with the 2D situation, where point-like artifacts (like noise) are smoothed, while edges in the image are preserved. A similar trajectory was already described in the context of *kt*-BLAST [16], where the third dimension was time instead of the *kz*-direction. This effect is illustrated in Figure 2. A stack of stars acquisition with these modified angles now benefits from the TV filtering in z-direction, which improves the overall artifact reduction capability of the ROF model.

2.2 A Fast Minimization Algorithm

In the following we derive the PDU algorithm for the ROF model [31] and present efficient step sizes for 3D volumes.

The dual variable $\mathbf{p}$ $(= \left(p^1, p^2, p^3\right)^T$ in 3D) for given primal value u is defined such that

$$|\nabla u| = \sup_{||\mathbf{p}|| \leq 1} \{\mathbf{p} \cdot \nabla u\} . \tag{2}$$
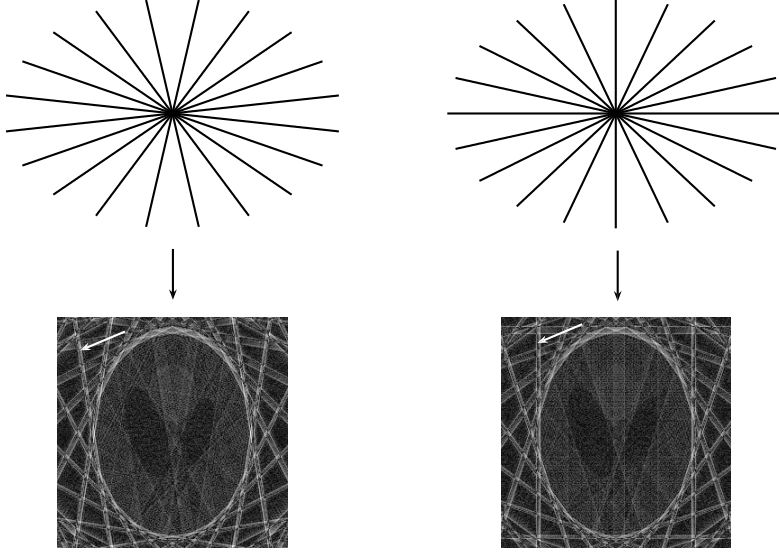
Fig. 2 Illustration of the effect of shifted projection angles on the streaking artifacts in the reconstructed image for a numerical phantom. The phantom was sampled using 10 radial projections, and the projection angles were shifted by $\frac{\pi}{2k}$ between the two reconstructions. This leads to changes in the structure of the streaking artifacts which can easily be seen by the orientation of the specific streak that is highlighted by the arrows.

By reformulating (1) using the dual variable $\mathbf{p}$ we arrive at the primal-dual formulation of the ROF model:

$$\min_{u} \sup_{\|\mathbf{p}\| \leq 1} \left\{ \int_{\Omega} \mathbf{p} \cdot \nabla u \, dx + \frac{\lambda}{2} \int_{\Omega} (u - f)^2 \, dx \right\} . \tag{3}$$

The problem is now a saddle point problem in two variables. This can be solved by alternatingly minimize with respect to u and maximize with respect to $\mathbf{p}$.

1. *Primal update:* For the primal update we differentiate (3) according to u, to get the following Euler-Lagrange (EL) equation:

$$-\operatorname{div} \mathbf{p} + \lambda (u - f) = 0 . \tag{4}$$

Performing a gradient descent update scheme leads to

$$u^{n+1} = u^n (1 - \tau_P^n) + \tau_P^n \left(f + \frac{1}{\lambda} \operatorname{div} \mathbf{p} \right) , \tag{5}$$

where τ_P^n denotes the step size for the primal update.

2. *Dual update:* Differentiating (3) with respect to $\mathbf{p}$ we get the following EL equation:

$$\nabla u + \mathbf{p}\alpha = \mathbf{0} , \tag{6}$$

where α is a Lagrange multiplier for the additional constraint $||\mathbf{p}|| \leq 1$. The result is a gradient ascent method with a subsequent re-projection to restrict the length of $\mathbf{p}$ to 1:

$$\mathbf{p}^{n+1} = \Pi_{B_0} \left(\mathbf{p}^n + \tau_D^n \nabla u \right) . \tag{7}$$

Here $B_0 = \{ p : ||p|| \leq 1 \}$ denotes the unit ball centered at the origin, and τ_D^n is the dual step size. The projection onto B_0 can be formulated pointwise as

$$\Pi_{B_0} (\mathbf{q}) = \frac{\mathbf{q}}{\max\{ 1, |\mathbf{q}| \}} . \tag{8}$$

These two steps are iterated until convergence. Similar to [31], in 3D the following step size scheme offers good results for all our testing data:

$$\begin{array}{l} \tau_D^n = 0.3 + 0.02n \\ \tau_P^n = \frac{1}{\tau_D^n} \left(\frac{1}{6} - \frac{5}{15+n} \right) . \end{array} \tag{9}$$

This choice of step sizes is the most crucial part of this minimization algorithm, as constant or poorly chosen step sizes lead to significantly higher convergence times. However, it needs to be pointed out that the algorithm converges to the same minimizer for every choice of step size which is smaller than a constant which only depends on the problem formulation. In particular, the given values are independent of the image to be reconstructed, and thus need not be adapted by the user.

The primal-dual gap [31] can be used as a convergence criterion. In order to evaluate this gap, the primal and dual energies have to be computed.

1. *Primal energy:* The primal energy can be calculated directly based on (1):

$$E_P = \int_\Omega |\nabla u| \, dx + \frac{\lambda}{2} \int_\Omega (u - f)^2 \, dx . \tag{10}$$

2. *Dual energy:* Using (4) we can reconstruct u from the dual variable $\mathbf{p}$:

$$u = \frac{1}{\lambda} \mathrm{div} \, \mathbf{p} + f . \tag{11}$$

Thus u can be eliminated from the primal-dual formulation in (3), and the dual energy is given by

$$E_D = \frac{1}{2\lambda} \int_\Omega (\mathrm{div} \, \mathbf{p})^2 \, dx + \int_\Omega f \mathrm{div} \, \mathbf{p} \, dx . \tag{12}$$

As the optimization scheme consists of a minimization and a maximization problem, E_P presents an upper bound and E_D presents a lower bound for the true minimum of the ROF model. The primal-dual gap is then defined as

$$G(u, \mathbf{p}) = E_P (u) - E_D (\mathbf{p}) . \tag{13}$$

In [32], it was shown that

$$||u - u^*||_2 \leq \sqrt{\frac{G(u, \mathbf{p})}{\lambda}} , \tag{14}$$

where u^* denotes the global minimizer. The primal-dual approach therefore delivers a suitable convergence criterion.

Figure 3 shows an representative example of the behavior of the primal-dual gap during the iteration. Here, the gap was calculated in each iteration for the purpose of illustration. Usually, to save computation time, we evaluate the gap only every $N = 50$ iterations.
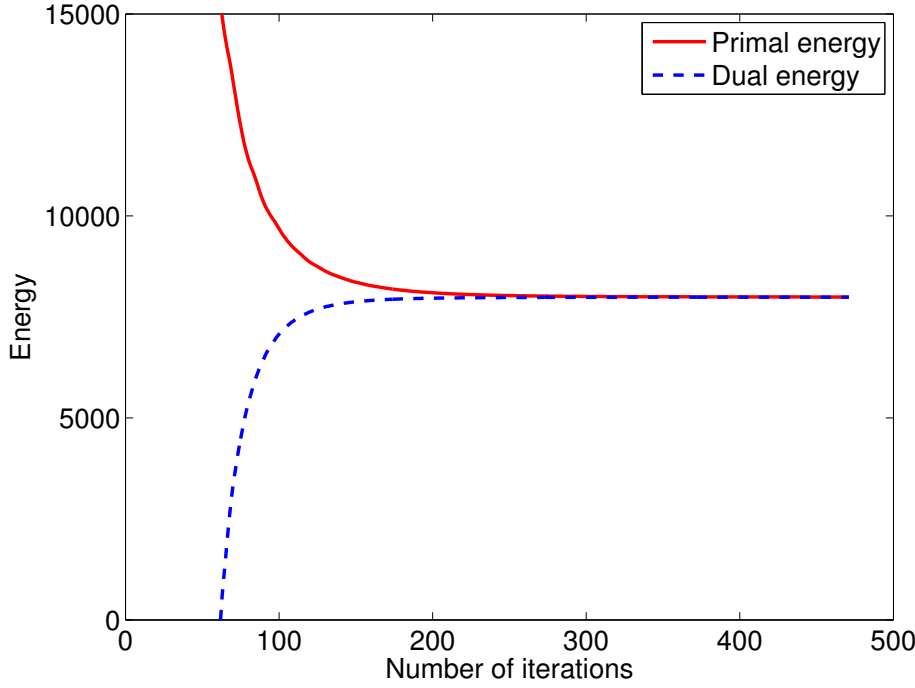
Fig. 3 Plot of the primal and dual energy during the iteration. The minimal functional value (1) always has to be smaller than the primal energy and greater than the dual energy. This plot corresponds to the 64 projections phantom data set in Figure 4.

2.3 GPU Implementation

Computing hardware have shown a clear tendency towards more and more parallelization over the last years. While modern CPUs already use 4 cores, modern GPUs like the Nvidia GeForce GTX 280 (Nvidia, Santa Clara, CA) utilize 240 cores. It should be noted that CPUs offer more programming flexibility, and GPUs use SIMD (single instruction, multiple data) architectures that require data-parallel algorithms. Additionally, GPUs utilize pipelining principles to optimize data throughput, the Nvidia GeForce GTX 280 offering a memory bandwidth of $141.7\,\mathrm{GB/s}$. To utilize this high bandwidth, an efficient memory management is a crucial part of GPU implementations.

To account for this change in computer hardware, special care has to be taken during choice and development of algorithms. Variational methods have an inherent parallelism, and are therefore perfectly suited for modern GPUs. Note that in every iteration of equations (5) and (7), each voxel only needs the values of its neighbors from the last iteration. Therefore, a voxel-wise parallelization can be achieved.

Our implementation was done using the CUDA [21] framework, which allows C-like programming on the GPU. For each voxel a single thread is created on the GPU, and synchronization is performed after each iteration. Special care was taken to maintain coalesced memory access during the whole computation. The GPU is organized in multiprocessors consisting of 8 single cores that have access to a fast local memory. This memory is orga-

nized in 16 banks which can be accessed simultaneously, and is as fast as registers when no bank conflicts occur. We make heavy use of this local memory by first loading a small patch from global memory, performing one iteration on this patch, and writing the patch back to global memory.

When implementing the algorithm from section 2.2, one has to consider the discrete nature of the data. This implies that we work on a cubic image domain $\Omega = [x_1, x_m] \times [y_1, y_n] \times [z_1, z_o]$. The discrete grid points are given by

$$(x_i, y_j, z_k) = (i\Delta x, j\Delta y, k\Delta z) \,, \tag{15}$$

where Δx, Δy and Δz denote the spatial discretization steps. Since voxel size is usually not isotropic in medical data sets, it is important to allow for non-uniform grid sizes.

The consistent discretization of the derivative operators is of great importance. Here, the discrete gradient operator is

$$(\nabla u)_{i,j,k} = \left(\delta_x^+ u_{i,j,k}, \delta_y^+ u_{i,j,k}, \delta_z^+ u_{i,j,k} \right)^T \,, \tag{16}$$

where the forward differences are defined as

$$
\begin{aligned}
\delta_x^+ u_{i,j,k} &= \begin{cases} \frac{u_{i+1,j,k} - u_{i,j,k}}{\Delta x} & \text{if } i < m \\ 0 & \text{if } i = m \end{cases} \\
\delta_y^+ u_{i,j,k} &= \begin{cases} \frac{u_{i,j+1,k} - u_{i,j,k}}{\Delta y} & \text{if } j < n \\ 0 & \text{if } j = n \end{cases} \\
\delta_z^+ u_{i,j,k} &= \begin{cases} \frac{u_{i,j,k+1} - u_{i,j,k}}{\Delta z} & \text{if } j < o \\ 0 & \text{if } j = o \,. \end{cases}
\end{aligned}
\tag{17}
$$

The discrete divergence operator must be constructed as the adjoint of the gradient:

$$(\text{div } \mathbf{p})_{i,j,k} = \delta_x^- p_{i,j,k}^1 + \delta_y^- p_{i,j,k}^2 + \delta_z^- p_{i,j,k}^3 \,, \tag{18}$$

where the backward differences are given by

$$
\begin{aligned}
\delta_x^- p_{i,j,k}^1 &= \begin{cases} \frac{p_{i,j,k}^1 - p_{i-1,j,k}^1}{\Delta x} & \text{if } i > 1 \\ p_{i,j,k}^1 & \text{if } i = 1 \end{cases} \\
\delta_y^- p_{i,j,k}^2 &= \begin{cases} \frac{p_{i,j,k}^2 - p_{i,j-1,k}^2}{\Delta y} & \text{if } j > 1 \\ p_{i,j,k}^2 & \text{if } j = 1 \end{cases} \\
\delta_z^- p_{i,j,k}^3 &= \begin{cases} \frac{p_{i,j,k}^3 - p_{i,j,k-1}^3}{\Delta z} & \text{if } k > 1 \\ p_{i,j,k}^3 & \text{if } k = 1 \,. \end{cases}
\end{aligned}
\tag{19}
$$

Referring to (14), we note that the primal-dual gap $G(u, p)$ can serve as an objective convergence criterion. To be independent of the size of the data, the input f is normalized between $[0, 1]$ and the following convergence criterion is used:

$$\sqrt{\frac{G(u, \mathbf{p})}{\lambda M^2}} < \zeta \,, \tag{20}$$

with $M = m \cdot n \cdot o$ the number of pixels and ζ the convergence threshold. As can be seen from (14), our convergence metric is an upper bound for the root mean squared error (RMSE) to the true global minimizer. We chose $\zeta = 10^{-6}$ throughout the experiments. Note that when

using 16 bit data, the gray value quantization step is 1.53×10^{-5}. Therefore our convergence threshold already delivers highly accurate results. For pure visual inspection, a higher convergence threshold could be chosen, which would further reduce computation times.

A software library that provides the described algorithm, as well as an interactive application for 3D data sets that was used throughout this paper, is available online at `http://www.gpu4vision.org`.

3 Materials and Methods

3.1 Phantom Measurements

The proposed radial stack of stars sequence with shifted projection angles was implemented on a clinical MR scanner (Siemens Magnetom TIM Trio, Erlangen, Germany).

An angiography phantom was constructed by inserting a plastic tube filled with with a Gd-DTPA (Magnevist, Schering AG) solute in a bottle of distilled water doped with MnCl2. The T_1 relaxation time of the tube is approximately 50 ms, which is comparable to a typical relaxation rate in the vessel system during the first passage of a contrast agent. The surrounding water in the bottle has a T_1 time of approximately 800 ms, comparable to the relaxation time of white matter in the brain. The phantom therefore represents the situation of a CE-MRA measurement of the brain vessels.

A 2D multi slice gradient echo sequence with the following sequence parameters, which ensured a strong T_1 contrast, was used to acquire the phantom images: TR = 12 ms, TE = 5 ms, FA = $60°$, matrix size (x,y) = 256×256, 11 slices with a slice thickness of 2.5 mm, imaging field of view FOV = 150 mm $\times$ 150 mm. All measurements were conducted at 3 T, using a transmit and receive birdcage resonator head coil. Measurements were performed with 64, 32 and 24 projections, resulting in undersampling factors of $R = 4$, $R = 8$ and $R \approx 10$ below the Nyquist limit in comparison to a fully sampled Cartesian data set. The corresponding MRI acquisition times are 8.4 s (64 projections), 4.2 s (32 projections) and 3.2 s (24 projections).

Raw data was exported from the scanner and offline image reconstruction was performed using a MATLAB (The MathWorks, Natick, MA) implementation of the non-uniform fast Fourier transform (NUFFT) [11]. Afterwards, these images were processed with the proposed 3D TV GPU method.

3.2 In vivo Angiography Measurements

The 3D TV GPU method was also evaluated with an in vivo data set, and the results were compared with reconstructions using alternative strategies. To assess image quality quantitatively, a fully sampled contrast enhanced MR angiography (CE-MRA) data set of the carotid arteries was acquired on a clinical MR scanner at 3 T (Siemens Magnetom TIM Trio, Erlangen, Germany) using a 3D FLASH sequence. Sequence parameters were repetition time TR = 3.74 ms, echo time TE = 1.48 ms, flip angle FA = $20°$, matrix size (x,y,z) = 448 $\times$ 352 $\times$ 40, voxel size $(\Delta x, \Delta y, \Delta z)$ = 0.55 mm $\times$ 0.55 mm $\times$ 0.70 mm. The data set was exported and retrospectively subsampled in the xy-plane to simulate an accelerated acquisition. The fully sampled data set served as the gold standard reference for quantitative evaluations. Acquisitions with 80 and 40 projections, corresponding to undersampling factors of $R = 5.6$ and $R = 11.2$ in comparison to a fully sampled Cartesian data set, were simulated. These

undersampling rates were chosen to illustrate situations when the TV reconstruction is able to eliminate almost all streaking artifacts (80 projections) and scenarios where the amount of undersampling is too high and residual artifacts remain after TV reconstruction (40 projections). For these sequence parameters, MRI acquisition times of the accelerated acquisitions would be 12.0 s (80 projections) and 6.0 s (40 projections). The MATLAB implementation of the non-uniform fast Fourier transform (NUFFT) [11] was again used during offline image reconstruction, and the images were then processed with the 3D TV filter.

To show the benefits of the proposed shifted spokes trajectory together with 3D TV, the results were compared to the application of the algorithm to a conventional not shifted stack of stars trajectory and to 2D slice by slice TV filtering. Additionally, TV reconstruction with a k-space data fidelity term [3] was implemented using the nonlinear conjugate gradient algorithm from [17]. The regularization parameter λ was chosen according to visual inspection of the reconstruction quality. In each experiment, we started with very low regularization and gradually increased it until artifacts started to disappear. This iterative procedure was stopped when the last streaking artifact was removed. As this could not be achieved for the 40 projections in vivo data set for some methods, the increase of the regularization was stopped as soon as pronounced image features were lost. Approximately 5-10 runs, depending on the data set, were performed in that way for each method.

Image quality was quantified by means of the root-mean-square (RMS) difference to the fully sampled data set, normalized by the RMS intensities of the fully sampled images. RMS differences were evaluated slice by slice. Additionally, mean value and standard deviation over all 40 slices were calculated.

3.3 Reconstruction Time and Convergence Behavior of the GPU Implementation

The proposed PDU 3D TV algorithm was implemented on the GPU. As the goal of this work was to evaluate the speedup that can be gained with the GPU implementation and not a comparison of reconstruction times for different algorithms, computation times were only evaluated for the proposed method. While the analysis of multiple methods is important in the case of image quality to show the benefits of 3D regularization, the most important test to evaluate speedup is a comparison to a fast C++ implementation of the same primal-dual algorithm on the CPU. Additionally, the PDU approach was compared with the fastest (to our knowledge) 3D implementation of the ROF model on the GPU [23], which uses Chambolle's projected gradient descent algorithm (CPG). This algorithm is known to have a fast convergence in the beginning, but to slow down towards the end of the optimization process. For that reason, both computation time and convergence behavior were investigated.

Experiments were performed on an Intel Core 2 Duo 6700, while the GPU implementation was tested on an Nvidia GeForce GTX 280 using CUDA 2.0. All CPU times are reported for computations on a single core, in order to allow accurate timing and to disregard any scheduling overhead. Of course, an implementation on multi-core architectures would be parallelized in practice (e.g., using OpenMP), yielding a higher performance corresponding to the number of cores.

4 Results

Figure 4 compares the results of conventional NUFFT reconstructions and the proposed TV method for the phantom experiments. It is not surprising that the conventional NUFFT re-

Table 1 Quantitative evaluation of the following reconstruction methods for the in vivo angiography data: Conventional NUFFT reconstruction with zero filling, 2D slice by slice ROF TV filtering, TV reconstruction with data fidelity in kspace [3], proposed 3D TV filtering using a conventional stack of stars trajectory, proposed 3D TV filtering using the shifted stack of stars trajectory that is introduced in this paper. Mean value and standard deviation of RMS differences (a.u.) to the fully sampled data set for 40 slices are displayed using the 80 and 40 projections subsampled data.

Data set	NUFFT	2D TV	2D kspace	3D TV	3D TV shift.
80 proj.	0.40 ± 0.03	0.25 ± 0.02	0.26 ± 0.02	0.26 ± 0.02	0.25 ± 0.02
40 proj.	0.64 ± 0.06	0.39 ± 0.03	0.35 ± 0.02	0.34 ± 0.01	0.31 ± 0.01

constructions suffer from characteristic streaking artifacts which become increasingly worse as the number of projections is reduced. Artifacts are efficiently reduced by application of the TV filter. Additionally, the scans of the phantom are significantly deteriorated by noise, which in the TV reconstructions is also reduced due to its inherent noise cancellation properties. The final regularization parameters were $\lambda_{64proj} = 10$, $\lambda_{32proj} = 7$ and $\lambda_{24proj} = 5$. Here, smaller parameters amount to stronger regularization (i.e., the influence of the data fidelity term is reduced, as can be seen in equation (3)).

Figure 5 shows the reconstruction results of all tested methods for the downsampled angiography data set with 80 and 40 projections. Similar to the phantom experiments, conventional NUFFT reconstructions show streaking artifacts. While all TV methods reduce artifacts, only the 3D TV method with shifted projections is nearly artifact free for the 80 projections data set. In contrast, residual artifacts can be seen in all results of the 40 projections data set, but best image quality is again achieved with the 3D TV method with shifted projections. The final regularization parameters were $\lambda_{80proj} = 0.03$ and $\lambda_{40proj} = 0.08$ for 2D ROF TV, $\lambda_{80proj} = 0.35$ and $\lambda_{40proj} = 0.45$ for the implementation with a k-space data fidelity term, $\lambda_{80proj} = 15$ and $\lambda_{40proj} = 10$ for 3D TV with conventional and shifted radial sampling. It must be pointed out that these parameters cannot be compared directly between the methods due to the different formulation. Particularly, for the 3D methods, smaller parameters amount to stronger regularization as noted. On the other hand, the parameter λ in the 2D implementations corresponds to $\frac{1}{\lambda}$ in the ROF formulation, so that larger values of λ indicate higher regularization. Additionally, the relation between L^2 and L^1 norms is different in 2D and 3D, and so the parameter value for a given balance changes with dimension.

The results of the quantitative analysis of image quality for all methods are displayed in Table 1. In the case of 80 projections, all TV reconstruction methods show comparable RMS differences which are significantly lower than for the conventional NUFFT reconstruction. This is also true for the 40 projections data set, but both conventional 3D TV and especially 3D TV with shifted projections show significantly lower RMS differences than the other methods.

A compilation of computation times of 3D PDU TV for all tested data sets can be found in Table 2. With the GPU implementation, a computation speed of 718 iterations per second for the phantom data set and 129 iterations per second for the in vivo data set was achieved. In comparison, the computation speed of the CPU implementation was 3.78 iterations per second for the phantom data set and 0.46 iterations per second for the in vivo data set. This corresponds to speedup factors of 190 (phantom data) and 280 (in vivo data) with the GPU. Note that the GPU scales much better with increasing size of the data set, as the parallel hardware is optimized for a high data throughput.

The convergence behavior of the proposed primal-dual approach to a GPU implementation of the CPG algorithm and our CPU implementation of the PDU algorithm is illustrated
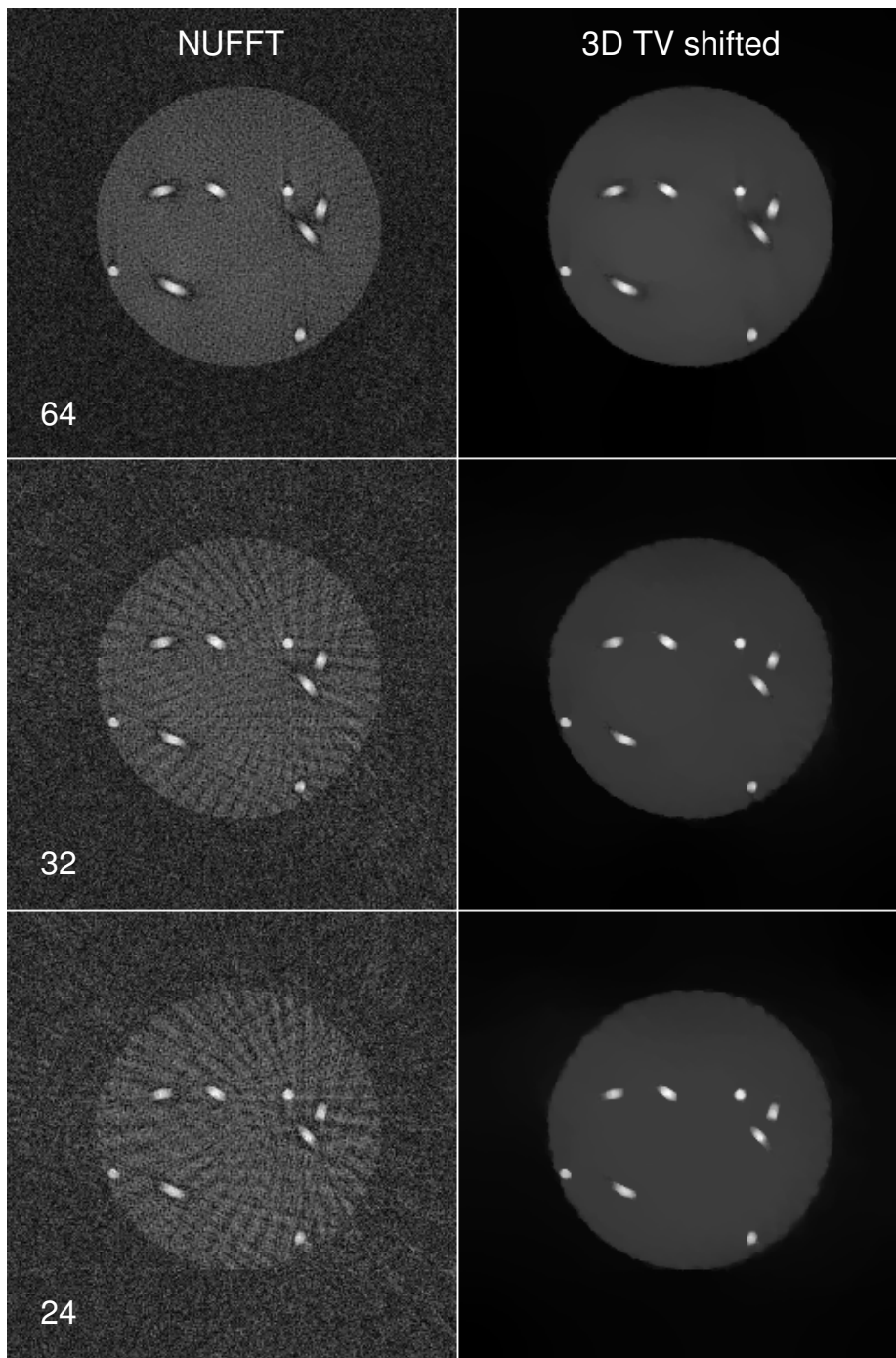
Fig. 4 A single slice from the data set of the phantom measurements using 64 (top row, $\lambda = 10$), 32 (middle row, $\lambda = 7$) and 24 (bottom row, $\lambda = 5$) shifted projections. Conventional NUFFT reconstruction (left) and reconstruction with the proposed 3D TV method.
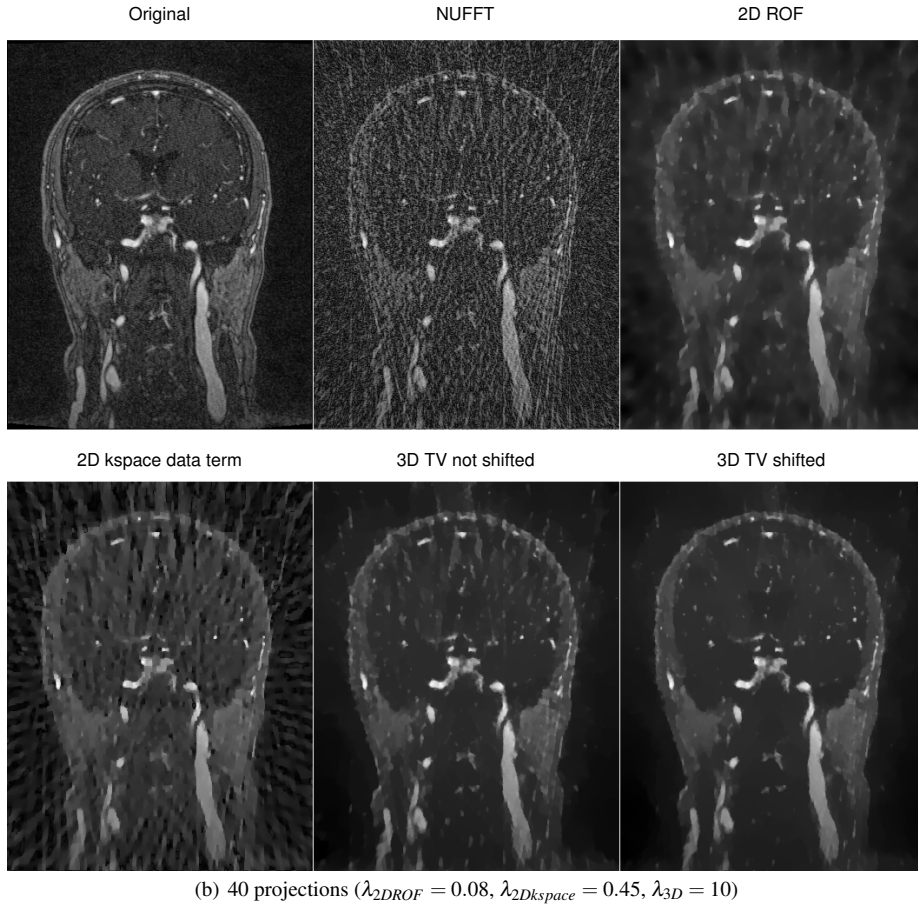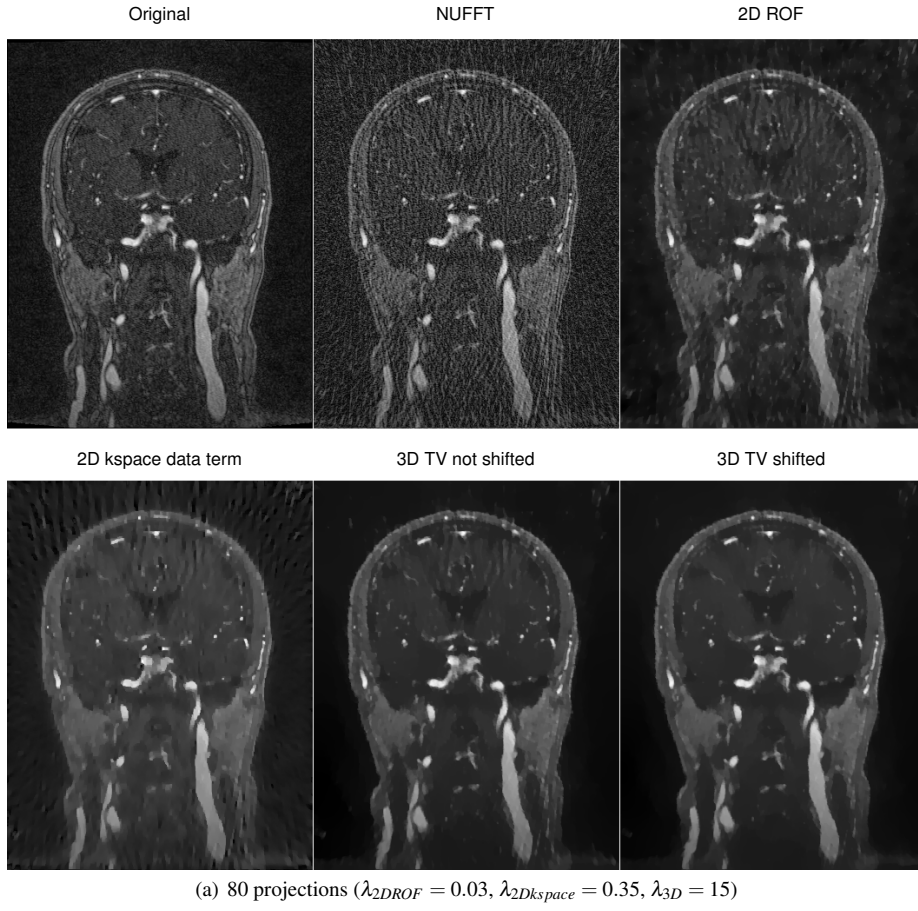
Original NUFFT 2D ROF

2D kspace data term 3D TV not shifted 3D TV shifted

(a) 80 projections ($\lambda_{2DROF} = 0.03$, $\lambda_{2Dkspace} = 0.35$, $\lambda_{3D} = 15$)

Original NUFFT 2D ROF

2D kspace data term 3D TV not shifted 3D TV shifted

(b) 40 projections ($\lambda_{2DROF} = 0.08$, $\lambda_{2Dkspace} = 0.45$, $\lambda_{3D} = 10$)

Fig. 5 A single slice from the angiography data set with 80 projections (a) and with 40 projections (b). Shown are original, fully sampled data set, conventional NUFFT reconstruction with zero filling, 2D slice by slice ROF TV filtering, TV reconstruction with data fidelity term in kspace, proposed 3D TV filtering using a conventional stack of stars trajectory, proposed 3D TV filtering using the shifted stack of stars trajectory

Table 2 Computation times of PDU 3D TV on the GPU for all tested data sets.

Data set	Size	Time (s)	Iterations
Phantom 64 projections	$256 \times 256 \times 11$	0.556	400
Phantom 32 projections	$256 \times 256 \times 11$	0.697	500
Phantom 24 projections	$256 \times 256 \times 11$	0.836	600
In vivo 80 projections	$448 \times 352 \times 40$	1.546	200
In vivo 40 projections	$448 \times 352 \times 40$	1.552	200
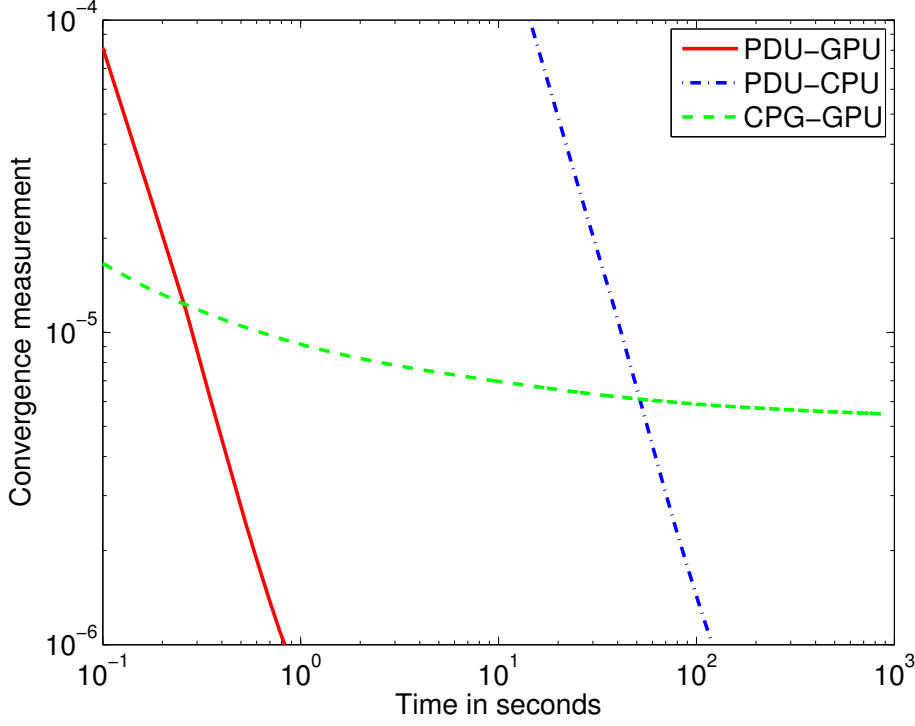
Fig. 6 Comparison of convergence behavior of the proposed GPU algorithm (PDU-GPU), a CPU implementation of the same algorithm (PDU-CPU) and a projected gradient descent algorithm on the GPU (CPG-GPU). These runs correspond to the 32 projections phantom data set in Figure 4. Shown is the primal dual gap metric given by the left hand side of (20).

in Figure 6, which plots the primal dual gap metric given by the left hand side of (20). One can clearly note that the PDU algorithm has a significantly faster convergence behavior. More importantly, the PDU algorithm achieves a significantly smaller primal-dual gap. In this example the CPG algorithm did not manage to fulfill our convergence criterion of $\zeta = 10^{-6}$. For the purpose of illustration, we show the phantom data set in Figure 6, but the results for the in vivo data are similar.

5 Discussion

Our reconstructions with 3D TV and the shifted spokes sampling pattern show excellent removal of undersampling artifacts even at high acceleration factors. Due to the nature of the ROF functional (1), vessels with their strong contrast are preserved because the L^2 norm in the data fidelity term makes the removal of structures contrast dependent. Current work is concerned with extending the proposed approach to L^1 data fidelity terms, which are contrast invariant and thus better suited to images with low contrast features. Of course it must be mentioned that the results from the experiments with the angiography phantom represent an optimal situation for our algorithm as the phantom only consists of piecewise constant areas. This explains why all artifacts could be eliminated even for an undersampling factor of $R \approx 10$, which was not possible for the in vivo angiography data set. Future work will be necessary to evaluate TV methods in clinical studies.

Visual inspection of the image quality for the in vivo data set showed that 3D TV, together with the shifted spokes acquisition, significantly improved removal of streaking artifacts for both subsampling factors considered. Quantitative analysis also resulted in significantly lower RMS differences for the 40 projections data set, while RMS differences were similar for all TV-based methods in the case of 80 projections. As the images clearly show different levels of artifact corruption, RMS difference cannot be considered an optimal metric to describe image quality of MR images. However, due to the lack of a better quality metric, it is usually used in the literature and hence is included here to facilitate comparison with other works.

As already mentioned in section 2, the biggest difference between our approach and the method recently described in [3] is that in our work, the data fidelity term in the ROF model is evaluated in image space. Whereas we use a two step procedure in which a Fourier transform is performed first, and TV based artifact removal is applied as a second step, the algorithm in the cited reference incorporates the TV penalty directly during the reconstruction process. As a consequence, the computation times of the two methods cannot be compared directly. While our approach has the advantage that the problem can easily be reformulated in the highly parallelized way necessary for implementation on the GPU architecture, it must be noted that it limits the algorithm to applications such as angiography where the structures of interest have a high contrast-to-noise-ratio. On the other hand, angiography is an application which benefits significantly from the possibility of real-time imaging. It should further be mentioned (as it is in [3] and [4]) that the method in [3] is also limited to specific applications due to the nature of the TV constraint. For the in vivo angiography data set that was investigated in this paper, no improvement in image quality was observed with a data fidelity term in kspace over conventional 2D ROF TV.

One important point concerning the comparison of different reconstruction strategies in this paper is that the regularization parameter was chosen based on visual inspection of the image quality. This was performed individually for each different method to compare best-case results for each method. In the absence of robust and objective metrics for medical image quality which could be used as a basis for automatic regularization parameter choice rules, visual inspection by medical experts is still the most sensible criterion. It is important to note that for our GPU-based method, this becomes a practicable approach, since this step can be performed with a tool which allows interactive adjustment of the parameter and continuously displays the effect on the image quality. This means that regularization can be adjusted similar to the way windowing is performed at the moment. In this way, selection of an optimal parameter could be done in less than 20 seconds for the in vivo data set. In comparison, each run of the CPU implementation takes approximately 7 minutes, which is

clearly not suitable for practical use. Of course, while visual inspection was our method of choice for this work, the GPU speedup can similarly be exploited in other parameter choice strategies that require multiple runs of the same optimization problem, such as heuristics of L-curve type and discrepancy principles [10], [20].

Finally, while 3D regularization was studied in this work, undersampling was only applied in the xy plane. Goals of future work will include the application to data sets where additional acceleration is included in the z-direction or for full 3D projection acquisition strategies like VIPR [1] where the aliasing pattern becomes more complicated.

Concerning the computation time, it was shown that a primal-dual approach with an appropriate step size choice outperforms current dual approaches for the ROF model, both in convergence time and exactness of the solution.

On current graphics hardware with 1 GB of memory, data sets of a maximum size of $512 \times 512 \times 204$ can be calculated at once, which is sufficient for typical MRI data sets. For bigger data sets, Nvidia Tesla cards could be used or computation could be split up on multiple GPUs.

With computation times of approximately 1.5 s for a $448 \times 352 \times 40$ data set, the GPU implementation allows TV filtering that is faster than the corresponding data acquisition times (12 s and 6 s for 80 and 40 projections, respectively), even in the case that multiple runs have to be performed to tune the regularization parameter. Therefore, the application of TV based artifact removal is no longer the time limiting step in the imaging chain. The goal of future work is to connect dedicated GPU computation hardware directly to the MR scanner, as was already described in the context of parallel imaging [26]. This will allow interactive TV artifact elimination already during data acquisition.

6 Conclusions

The results from this work show that the extension of TV filtering to 3D in combination with the proposed shifted stack of stars sampling trajectory leads to pronounced improvements in image quality for CE-MRA data. However, it must be noted that it cannot be expected that these results will generalize for arbitrary types of scans. With GPU implementations, TV computation times can be accelerated significantly and even allow interactive elimination of the regularization parameter for 3D TV. We believe that this can pave the way for TV based regularization strategies, currently promising research topics, to become powerful tools in daily clinical practice.

References

1. Barger, A.V., Block, W.F., Toropov, Y., Grist, T.M., Mistretta, C.A.: Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3d projection trajectory. Magn Reson Med **48**(2), 297–305 (2002)
2. Bernstein, M.A., King, K.F., Zhou, X.J.: Handbook of MRI Pulse Sequences. Academic Press (2004)

3. Block, K.T., Uecker, M., Frahm, J.: Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med **57**(6), 1086–1098 (2007)
4. Block, T.: Advanced methods for radial data sampling in MRI. Ph.D. thesis, Georg-August-Universitaet Goettingen (2008)
5. Carter, J.: Dual methods for total variation-based image restoration. Ph.D. thesis, UCLA, Los Angeles, CA, USA (2001)
6. Chambolle, A.: An algorithm for total variation minimizations and applications. Journal of Math. Imaging and Vision **20**(1–2), 89–97 (2004)
7. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 136–152 (2005)
8. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Nummer. Math. **76**(167-188) (1997)
9. Chan, T., Golub, G., Mulet, P.: A nonlinear primal-dual method for total variation-based image restoration. SIAM Journal of Applied Mathematics **20**(6), 1964–1977 (1999)
10. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, *Mathematics and its Applications*, vol. 375. Kluwer Academic Publishers Group, Dordrecht (1996)
11. Fessler, J.A., Sutton, B.P.: Nonuniform fast Fourier transforms using min-max interpolation. IEEE Transactions on Signal Processing **51**(2), 560–574 (2003)
12. Goldfarb, D., Yin, W.: Parametric maximum flow algorithms for fast total variation minimization. Tech. rep., Rice University (2007)
13. Goldstein, T., Osher, S.: The split Bregman algorithm for L1 regularized problems. UCLA Computational and Applied Mathematics Reports **08**(29) (2008)
14. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med **47**(6), 1202–1210 (2002)
15. Hansen, M.S., Atkinson, D., Sorensen, T.S.: Cartesian SENSE and k-t SENSE reconstruction using commodity graphics hardware. Magn Reson Med **59**(3), 463–468 (2008)
16. Hansen, M.S., Baltes, C., Tsao, J., Kozerke, S., Pruessmann, K.P., Eggers, H.: k-t BLAST reconstruction from non-Cartesian k-t space sampling. Magn Reson Med **55**(1), 85–91 (2006)
17. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med **58**(6), 1182–1195 (2007)
18. Madore, B., Glover, G.H., Pelc, N.J.: Unaliasing by fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. Magn Reson Med **42**(5), 813–828 (1999)
19. Mistretta, C.A., Wieben, O., Velikina, J., Block, W., Perry, J., Wu, Y., Johnson, K., Wu, Y.: Highly constrained backprojection for time-resolved MRI. Magn Reson Med **55**(1), 30–40 (2006)
20. Morozov, V.A.: On the solution of functional equations by the method of regularization. Soviet Math. Dokl. **7**, 414–417 (1966)
21. NVIDIA: NVIDIA CUDA Programming Guide 2.0. NVIDIA Cooperation (2008)
22. Peters, D.C., Korosec, F.R., Grist, T.M., Block, W.F., Holden, J.E., Vigen, K.K., Mistretta, C.A.: Undersampled projection reconstruction applied to MR angiography. Magn Reson Med **43**(1), 91–101 (2000)
23. Pock, T., Unger, M., Cremers, D., Bischof, H.: Fast and exact solution of total variation models on the GPU. In: CVPR Workshop on Visual Computer Vision on GPU's. Anchorage, Alaska, USA (2008)
24. Popov, L.D.: A modification of the Arrow-Hurwicz method for search of saddle points. Mathematical Notes **28**(5), 845–848 (1980)
25. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: SENSE: sensitivity encoding for fast MRI. Magn Reson Med **42**(5), 952–962 (1999)
26. Roujol, S., de Senneville, B.D., Vahala, E., Srensen, T.S., Moonen, C., Ries, M.: Online real-time reconstruction of adaptive tsense with commodity cpu/gpu hardware. Magn Reson Med **62**(6), 1658–1664 (2009)
27. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(1-4), 259–268 (1992)
28. Sorensen, T.S., Schaeffter, T., Noe, K.O., Hansen, M.S.: Accelerating the nonequispaced fast Fourier transform on commodity graphics hardware. IEEE Transactions on Medical Imaging **27**(4), 538–547 (2008)
29. Tsao, J., Boesiger, P., Pruessmann, K.P.: k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations. Magn Reson Med **50**(5), 1031–1042 (2003)
30. Vogel, C., Oman, M.: Iteration methods for total variation denoising. SIAM Journal of Applied Mathematics. **17**, 227–238 (1996)
31. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report 08-34 (2008)
32. Zhu, M., Wright, S.J., Chan, T.F.: Duality-based algorithms for total variation image restoration. UCLA CAM Report 08-33 (2008)