SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

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

Florian Knoll Markus Unger

Christian Clason Clemens Diwoky
Thomas Pock Rudolf Stollberger

SFB-Report No. 2009–022 May 2009

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land
Steiermark

Stadt GRAZ Wissenschaft

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

Florian Knoll[*,a,1], Markus Unger[b], Clemens Diwoky[a], Christian Clason[c], Thomas Pock[b], Rudolf Stollberger[a]

[a] *Institute of Medical Engineering, Graz University of Technology, Kronesgasse 5/II, A-8010 Graz, Austria*
[b] *Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, A-8010 Graz, Austria*
[c] *Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria*

Abstract

Subsampling of radially encoded Magnetic Resonance Imaging (MRI) acquisitions in combination with L1 penalized iterative reconstruction methods opened a door to significantly increase the imaging speed in MRI which is crucial for many important clinical applications. In particular, these strategies have great potential to acquire contrast enhanced MR angiography data sets with high spatial and temporal resolution. It has been shown recently that image reconstruction with a Total Variation (TV) constraint efficiently reduces undersampling artifacts, but the drawback of the method is the long reconstruction time. This paper demonstrates that by implementing a primal-dual method on a modern graphic processing unit (GPU), TV computation times can be achieved which allow image reconstruction in a few seconds. This is even possible for large volume data sets where the TV constraint is evaluated in 3D. Together with a special radial sampling pattern, this improves image quality of the reconstructions significantly. Results from phantom measurements and in vivo angiography data sets are presented, which show excellent removal of streaking artifacts even for highly undersampled data sets.

Key words: MRI, Angiography, 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. Therefore, 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 as this violates the Nyquist criterion, it leads to artifacts in the reconstructed images. There are different strategies to mitigate these artifacts like parallel imaging [22, 13] or methods that exploit spatio-temporal correlations [16, 25, 17].

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 [15, 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 the major drawback is their long computation time. While this is not a severe limitation during research, it is currently impossible to use them in daily clinical practice.

Hansen et al. [14] and Sorensen et al. [24] showed 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 primal-dual based TV regularization strategies can

[*] Corresponding author

Email addresses: `florian.knoll@tugraz.at` (Florian Knoll), `unger@icg.tu-graz.ac.at` (Markus Unger), `clemens.diwoky@tugraz.at` (Clemens Diwoky), `christian.clason@uni-graz.at` (Christian Clason), `pock@icg.tu-graz.ac.at` (Thomas Pock), `rudolf.stollberger@tugraz.at` (Rudolf Stollberger)

[1] tel.: +43 316 873 5375, fax: +43 316 873 7890

be solved on the GPU, with computation times that allow online interactive elimination of undersampling artifacts. This is especially important for the determination of the regularization parameter because an optimal choice usually depends on patient-specific conditions like the geometry of the imaged anatomy. Therefore predefined settings may not deliver optimal results. It is also shown in this paper that by using a new type of radial sampling, 3D a priori information can be included in the reconstruction process, which significantly improves image quality. While this leads to an even higher computational complexity during the reconstruction, it becomes feasible with the GPU implementation.

In combination with highly undersampled radial trajectories which dramatically reduce data acquisition time, it opens new perspectives to perform real-time magnetic resonance imaging.

2. Theory

2.1. Undersampled Radial Imaging and Total Variation

It is well known [2] that $\frac{\pi}{2} \cdot n$ radial projections have to be sampled to obtain a fully sampled radial data set if an $n \times n$ image matrix has to be reconstructed. If the number of projections is reduced, this 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 compareable with compressed sensing approaches [15]. In contrast, the original TV approach is used in our approach. TV based minimization problems were originally designed for elimination of Gaussian noise, and were first used by Rudin, Osher and Fatemi [23] in 1992. The denoising model using a TV regularization together with a L^2 data term is therefore often referred to as the ROF model. It is defined as the following minimization problem:

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

where f is the original image data that contains the streaking artifacts due to subsampling, u is the reconstructed image and Ω is the image domain. The free parameter λ controls the amount of regularization. The L^2 norm of the data fidelity term makes the removal of structures contrast dependent. While low contrast regions are removed, strong contrast regions of the same size are kept. The L^1 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 where a Fourier transform is applied to the MRI raw data first, and the TV method is then applied to the reconstructed images which are corrupted by aliasing artifacts. In contrast, Block et al. use a single step procedure where the TV based regularization is integrated in the image reconstruction process.

As the functional defined in (1) is convex, the global optimum can be calculated. The minimization of the ROF model is a well studied problem. In the original formulation the model was solved using explicit time marching [23]. Other methods linearize the Euler-Lagrange equation [26], [8]. Duality based methods showed great increases in 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 [27] and [28] which were already used to find saddle points by Popov in [21]. In [12], a Split Bregman algorithm is used for minimization. Also discrete methods using graph cuts can be used to solve the ROF model [11]. As shown in [20], continuous methods have the advantage of their inherent parallelization potential, a low memory consumption and no discretization artifacts.

Recently, continuous methods have become real-time applicable for large 3D data sets by implementing them on the GPU [20]. It is shown in this paper that by adapting the PDU approach from [27] to 3D and by implementing it on the GPU, even large volume data sets can be calculated 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 slices. 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 [19] 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 slices, which creates the same streaking artifacts, and therefore does not benefit from the TV constraint in z-direction. But if the projection angles of adjacent slices are shifted by $\frac{\pi}{2k}$, when k is the number of projections (see Figure 1), the aliasing pattern in adjacent slices gets a different structure. This effect is illustrated in Figure 2. A stack of stars acquisition with these modified angles now benefits from the TV regularization in z-direction, which improves the overall artifact reductions performance of the ROF model.

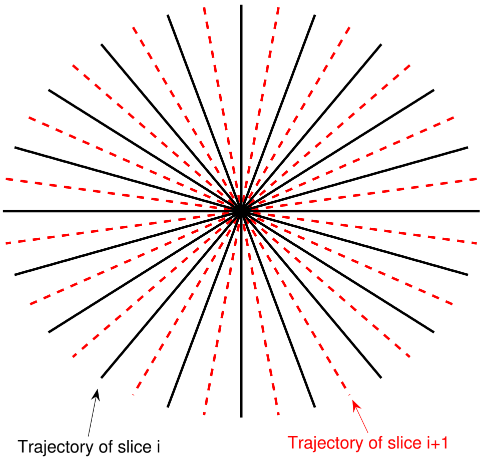

Figure 1: Stack of stars trajectory with shifted projection angles, 10 radial projections are displayed.

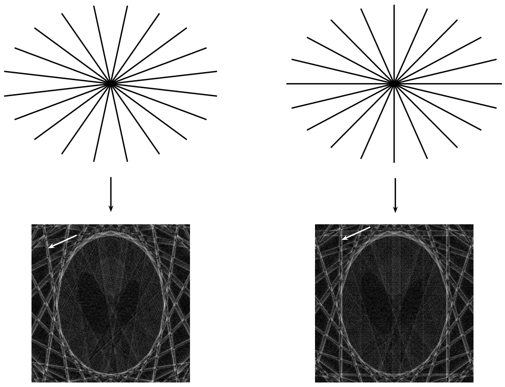

Figure 2: Illustration of the effect of shifted projection angles on the streaking artifacts in the reconstructed image for a numerical phantom. The phantom was sampled using 10 radial projections, and the projection angles were shifted by $\frac{\pi}{2k}$ between the two reconstructions. This leads 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.

2.2. A fast Minimization Algorithm

In the following we derive the PDU algorithm for the ROF model [27] and present efficient step widths for 3D volumes.

The dual variable $\boldsymbol{p}$ (in 3D $\boldsymbol{p} = \left(p^1, p^2, p^3\right)^T$) is defined such that

$$|\nabla u| = \sup_{\|\boldsymbol{p}\| \leq 1} \{\boldsymbol{p} \cdot \nabla u\} \ . \tag{2}$$

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

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

The problem is now an optimization problem in two variables. Thus one has to do alternating minimization with respect to u and $\boldsymbol{p}$.

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

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

   Performing a gradient descent update scheme leads to

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

   where τ_P^n denotes the step width for the primal update.
2. *Dual update:* Differentiating (3) with respect to $\boldsymbol{p}$ we get the following EL equation:

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

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

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

   Here B_0 denotes a d-dimensional unit ball centered at the origin, and τ_D^n is the dual step width. The re-projection onto B_0 can be formulated as

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

These two steps are iterated until convergence. Similar to [27], in 3D the following step width scheme offers good results for all our testing data:

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

The step width scheme is the most crucial part of this minimization algorithm, as constant or poorly chosen step widths lead to significantly higher convergence times. As a convergence criterion, the primal-dual gap [27] is taken into account. Therefore the primal and dual energies have to be evaluated.

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

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

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

$$u = \frac{1}{\lambda}\texttt{div}\ \boldsymbol{p} + f\ . \qquad (11)$$

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

$$E_D = \frac{1}{2\lambda} \int_\Omega (\texttt{div}\ \boldsymbol{p})^2\, dx + \int_\Omega f\texttt{div}\ \boldsymbol{p}\, dx\ . \quad (12)$$

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

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

In [28], it was shown that

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

Here u^* denotes the globally optimal solution. Thus the primal-dual approach delivers a reasonable convergence criterion.

In Figure 3 an exemplary primal-dual gap is shown. For illustration the gap was calculated in each iteration. Usually, to save computation time, we evaluate the gap only every $N = 50$ iterations.

2.3. GPU Implementation

Recent computing hardware showed a clear tendency towards more and more parallelization over the last years. While CPUs already use up to 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 offers a memory bandwidth of 141.7 GB/sec. To utilize this high bandwidth, an efficient memory management is a crucial part of GPU implementations.

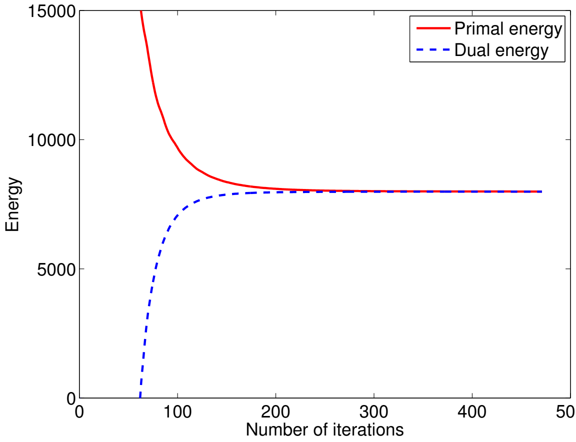

Figure 3: The primal-dual gap delivers a convergence criterion for the algorithm. The globally optimal solution always has to be smaller than the primal energy and greater than the dual energy. This run corresponds to the 64 projections phantom data set in Figure 5.

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 each iteration of equations 5 and 7, voxels only need the values of their neighbours from the last iteration. Therefore a voxel-wise parallelisation can be acchieved.

Our implementation was performed using the CUDA [18] framework that allows C-like programming on the GPU. For each voxel a single thread is created on the GPU, while after each iteration, synchronization is performed. 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 organized 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 locations of the grid are given by

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

where Δx, Δy and Δz denote the spatial discretization steps, that are specially important when working with medical data sets as the voxel size is usually not isotropic.

The discretization of the derivative operators is of great importance. In the discrete setting the gradient operator becomes:

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

The forward differences are defined as

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

The approximation of the divergence operator can be done in a similar manner:

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

where the backward differences are given as

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

The upper boundary for the root squared error in (14) delivers a well founded convergence criterion. To be independent of the size of the data, the following automatic convergence criterion was used:

$$\frac{\|u - u^*\|_2}{M} \leq \sqrt{\frac{G\left(u, \boldsymbol{p}\right)}{\lambda M^2}} < \zeta \ , \tag{20}$$

with $M = m \cdot n \cdot o$ the number of pixels, and ζ the convergence threshold. Our convergence measurement is an upper boundary for the root mean squared error (RMSE) to the true global optimum. We chose $\zeta = 10^{-6}$ throughout the experiments. For the calculation our data is normalized between [0, 1]. Note that when using 16bit data the gray value quantization step is $1.53 \cdot 10^{-5}$. Therefore our convergence threshold already delivers highly accurate results. For pure visual inspection a higher convergence threshold could be chosen, as this would result in faster convergence times.

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

3. 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 was approximately 50ms, 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 had a T_1 time of approximately 800ms which is 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 = 12ms, TE = 5ms, FA = 60°, matrix size (x,y) = 256 × 256, 11 slices with a slice thickness of 2.5mm, imaging field of view FOV = 150mm × 150mm. A transmit and receive birdcage resonator head coil was used. All measurements were conducted at 3T. Measurements were performed with 64, 32 and 24 projections. This results in undersampling factors of $R \approx 6$, $R \approx 12$ and $R \approx 16$ below the Nyquist limit in comparison to a fully sampled radial data set. The corresponding MRI acquisition times are 8.4s (64 projections), 4.2s (32 projections) and 3.2s (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) [10]. 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 different reconstruction 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 3T (Siemens Magnetom TIM Trio, Erlangen, Germany) using a 3D FLASH sequence. Sequence parameters were repetition time TR = 3.74ms, echo time TE = 1.48ms, flip angle FA = 20°, matrix size (x,y,z) = 448 × 352 × 40, voxel size (Δx,Δy,Δz) = 0.55mm × 0.55mm × 0.70mm. A single slice of the data set can be seen in Figure 4.

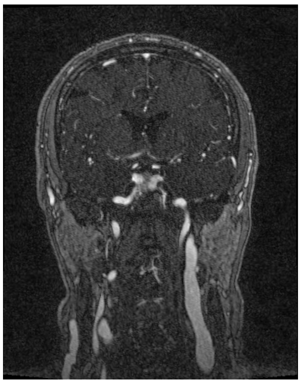

Figure 4: A single slice from the in vivo angiography data set that was used for the downsampling experiments.

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 $R \approx 9$ and $R \approx 18$ in comparison to a fully sampled radial data set, were obtained. These undersampling rates were chosen to illustrate situations when the TV reconstruction is able to eliminate almost all streaking artifacts ($R \approx 9$), in contrast to scenarios where the amount of undersampling is too high and residual artifacts remain after TV reconstruction ($R \approx 18$). For these sequence parameters, MRI acquisition times of the accelerated acquisitions would be 12.0s (80 projections) and 6.0s (40 projections). The Matlab implementation of the non-uniform fast fourier transform (NUFFT) [10] 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, the TV method that was proposed by Block et al. in [3] was implemented by using the nonlinear conjugate gradient algorithm from [15]. All regularization parameters were chosen according to visual inspection of the reconstruction quality.

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, 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 speedups 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 experiment to evaluate speedups 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 to our knowledge fastest 3D implementation of the ROF model on the GPU [20], where Chambolle's projected gradient descent algorithm (CPG) was used. CPG is known to have a good convergence in the beginning, but gets slow towards the end of the optimization process. Therefore both computation time and convergence behavior were investigated.

Experiments were performed on an Intel Core 2 Duo 6700, using only a single core. For the GPU implementations a Nvidia GeForce GTX 280 was used together with CUDA 2.0.

4. Results

4.1. Phantom Measurements

Figure 5 compares the results of conventional NUFFT reconstructions and the proposed TV method for the phantom experiments. It is not surprising that the conventional NUFFT reconstructions 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 is also reduced due to the inherent noise cancellation properties of the TV method.

4.2. In vivo Angiography Measurements

Figure 6 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 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

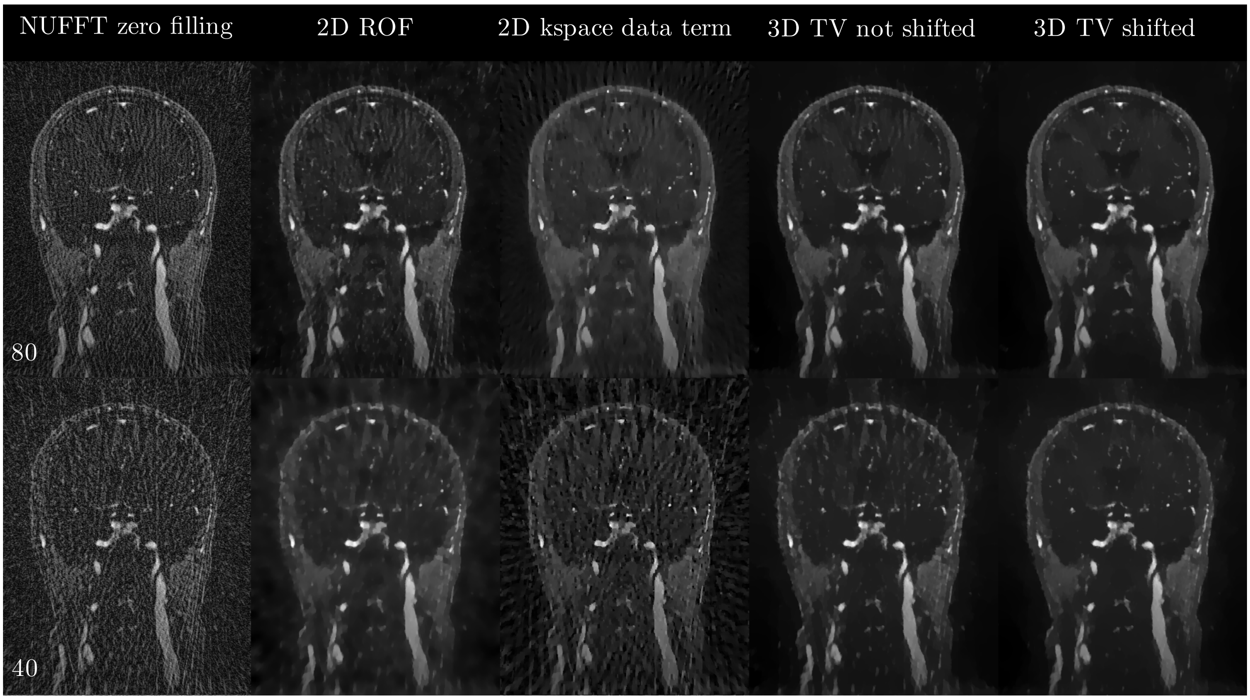

Figure 6: A single slice from the angiography data set with 80 projections (top row) and with 40 projections (bottom row). Conventional NUFFT reconstruction with zero filling, 2D slice by slice ROF TV filtering, TV reconstruction with data fidelity term in kspace (as proposed by Block et al in [3]), proposed 3D ROF TV filtering using a conventional stack of stars trajectory, proposed 3D ROF TV filtering using the shifted stack of stars trajectory that is introduced in this paper.

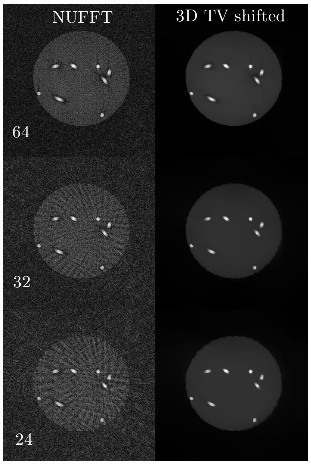

Figure 5: A single slice from the data set of the phantom measurements using 64 (top row), 32 (middle row) and 24 (bottom row) shifted projections. Conventional NUFFT reconstruction (left) and reconstruction with the proposed 3D TV method.

again achieved with the 3D TV method with shifted projections.

The results of the quantitative analysis of image quality for all methods are displayed in Table 1. All reconstruction methods show comparable RMS differences which are significantly better than the conventional NUFFT reconstruction in the case of 80 projections. 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.

4.3. Reconstruction Time

A compilation of computation times of 3D PDU TV for all tested data sets can be found in Table 2.

In Table 3, the performance of 3D TV CPU and GPU implementations are compared. The number of iterations does not depend on the number of projections, but only on the number of voxels in the final data set. With the chosen data sets, speedups of up to 280 can be gained. 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.

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 ROF TV filtering using a conventional stack of stars trajectory, proposed 3D ROF 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	0.40 ± 0.02	0.34 ± 0.01	0.31 ± 0.01

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

Data set	Time (s)	Iterations
Phantom 64 projections	0.556	400
Phantom 32 projections	0.697	500
Phantom 24 projections	0.836	600
In vivo 80 projections	1.546	200
In vivo 40 projections	1.552	200

Table 3: Computation speed of the PDU 3D TV algorithm: Comparison between implementations on GPU and CPU. The iterations per second represent the average of the experiments with different subsampling.

Data set	Size	Iterations per s		Speedup
		GPU	CPU	
Phantom	256 × 256 × 11	718	3.78	190
In vivo	448 × 352 × 40	129	0.461	280

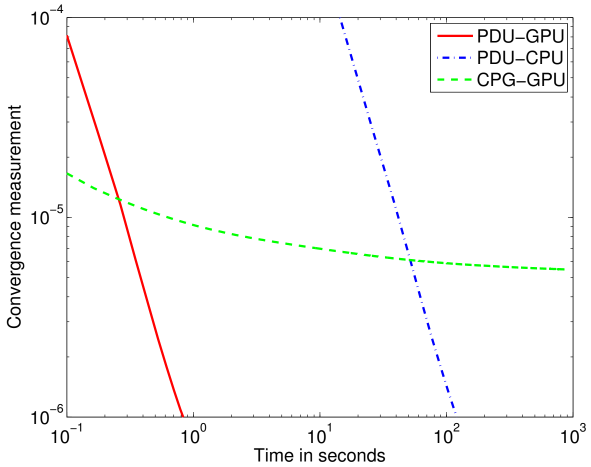

Figure 7: Comparison of convergence behavior of the proposed GPU algorithm (PDU-GPU), a CPU implementation of the same algorithm (PDU-CPU) and a projected gradient descent algorithm on the GPU (CPG-GPU). These runs correspond to the 32 projections phantom data set in Figure 5.

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 in Figure 7, which depicts the evolution of the upper boundary of the RMSE to the globally optimal solution. One can clearly note that the PDU algorithm has a significantly faster convergence behavior. And even more important, the PDU algorithm delivers a more exact solution of the energy. In this example the CPG algorithm did not manage to fulfill our convergence criterion of $\zeta = 10^{-6}$. In Figure 7, we chose the phantom data set as an example, but the in vivo data deliver similar results.

5. Discussion

5.1. Image Quality of the Reconstructions

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. 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 areas which are piecewise constant. This explains why all artifacts could be eliminated even for an undersampling factor of $R \approx 16$ 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 index, 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 of our approach to the method recently described in [3] is that is that in our work the data fidelity term in the ROF model is evaluated in image space. Where we use a two step procedure in which a Fourier transform is performed first, and TV regularization is applied as a second step, the algorithm in the cited reference includes 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 a highly parallelized way which is needed for implementation on the GPU architecture, it must be noted that it limits the algorithm to applications where the structures of interest have a high contrast-to-noise-ratio, like in angiography. On the other hand, angiography is an application which benefits significantly from the possibility of real-time imaging. It should furthermore 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. It is important to note that for the GPU-based methods, this becomes a practicable approach, since this step can be performed with an interactive tool which allows adjustments of the parameter and displays the effect on the image quality online and in real-time. 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 (and now feasible) criterion.

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.

5.2. Reconstruction Time

It was shown that a primal-dual approach, utilizing a reasonable time stepping scheme, outperforms current dual approaches to solve the ROF model, both in convergence time and exactness of the solution.

On current graphics hardware with 1GB 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. Alternatively, computation could be split up on multiple GPUs.

In conclusion, with computation times of approximately 1.5s for a $448 \times 352 \times 40$ data set, the GPU implementation facilitates TV image reconstruction that is faster than the corresponding data acquisition times (12s and 6s for 80 and 40 projections, respectively). Therefore, image reconstruction is no longer the time limiting step in the imaging chain. We believe that this can pave the way for TV reconstruction strategies, currently promising research topics, to become powerful tools in daily clinical practice.

Acknowledgments

This work was funded by the Austrian Science Fond under grant SFB F3209-18.

The authors would also like to thank Dr. S. Keeling of the Institute for Mathematics and Scientific Computing (University of Graz, Austria) for helpful discussions concerning the TV method, and B. Neumayer of the Institute of Medical Engineering (Graz University of Technology, Austria) for his help with the phantom measurements.

References

[1] Barger, A. V., Block, W. F., Toropov, Y., Grist, T. M., Mistretta, C. A., Aug 2002. Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3d projection trajectory. Magn Reson Med 48 (2), 297–305.

[2] Bernstein, M. A., King, K. F., Zhou, X. J., September 2004. Handbook of MRI Pulse Sequences. Academic Press.

[3] Block, K. T., Uecker, M., Frahm, J., Jun 2007. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med 57 (6), 1086–1098.

[4] Block, T., 2008. Advanced methods for radial data sampling in MRI. Ph.D. thesis, Georg-August-Universitaet Goettingen.

[5] Carter, J., 2001. Dual methods for total variation-based image restoration. Ph.D. thesis, UCLA, Los Angeles, CA, USA.

[6] Chambolle, A., 2004. An algorithm for total variation minimizations and applications. Journal of Math. Imaging and Vision 20 (1–2), 89–97.

[7] Chambolle, A., 2005. Total variation minimization and a class of binary MRF models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition. pp. 136–152.

[8] Chambolle, A., Lions, P.-L., 1997. Image recovery via total variation minimization and related problems. Nummer. Math. 76 (167-188).

[9] Chan, T., Golub, G., Mulet, P., 1999. A nonlinear primal-dual method for total variation-based image restoration. SIAM Journal of Applied Mathematics 20 (6), 1964–1977.

[10] Fessler, J. A., Sutton, B. P., Feb. 2003. Nonuniform fast Fourier transforms using min-max interpolation. IEEE Transactions on Signal Processing 51 (2), 560–574.

[11] Goldfarb, D., Yin, W., 2007. Parametric maximum flow algorithms for fast total variation minimization. Tech. rep., Rice University.

[12] Goldstein, T., Osher, S., Jun. 2008. The split Bregman algorithm for L1 regularized problems. UCLA Computational and Applied Mathematics Reports 08 (29).

[13] Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., Jun 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47 (6), 1202–1210.

[14] Hansen, M. S., Atkinson, D., Sorensen, T. S., Mar 2008. Cartesian SENSE and k-t SENSE reconstruction using commodity graphics hardware. Magn Reson Med 59 (3), 463–468.

[15] Lustig, M., Donoho, D., Pauly, J. M., Dec 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med 58 (6), 1182–1195.

[16] Madore, B., Glover, G. H., Pelc, N. J., Nov 1999. Unaliasing by fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. Magn Reson Med 42 (5), 813–828.

[17] Mistretta, C. A., Wieben, O., Velikina, J., Block, W., Perry, J., Wu, Y., Johnson, K., Wu, Y., Jan 2006. Highly constrained backprojection for time-resolved MRI. Magn Reson Med 55 (1), 30–40.

[18] NVIDIA, 2008. NVIDIA CUDA Programming Guide 2.0. NVIDIA Cooperation.

[19] Peters, D. C., Korosec, F. R., Grist, T. M., Block, W. F., Holden, J. E., Vigen, K. K., Mistretta, C. A., Jan 2000. Undersampled projection reconstruction applied to MR angiography. Magn Reson Med 43 (1), 91–101.

[20] Pock, T., Unger, M., Cremers, D., Bischof, H., Jun. 2008. Fast and exact solution of total variation models on the GPU. In: CVPR Workshop on Visual Computer Vision on GPU's. Anchorage, Alaska, USA.

[21] Popov, L. D., Nov. 1980. A modification of the Arrow-Hurwicz method for search of saddle points. Mathematical Notes 28 (5), 845–848.

[22] Pruessmann, K. P., Weiger, M., Scheidegger, M. B., Boesiger, P., Nov 1999. SENSE: sensitivity encoding for fast MRI. Magn Reson Med 42 (5), 952–962.

[23] Rudin, L. I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Phys. D 60 (1-4), 259–268.

[24] Sorensen, T. S., Schaeffter, T., Noe, K. O., Hansen, M. S., April 2008. Accelerating the nonequispaced fast Fourier transform on commodity graphics hardware. IEEE Transactions on Medical Imaging 27 (4), 538–547.

[25] Tsao, J., Boesiger, P., Pruessmann, K. P., Nov 2003. k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations. Magn Reson Med 50 (5), 1031–1042.

[26] Vogel, C., Oman, M., 1996. Iteration methods for total variation denoising. SIAM Journal of Applied Mathematics. 17, 227–238.

[27] Zhu, M., Chan, T., 2008. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report 08-34.

[28] Zhu, M., Wright, S. J., Chan, T. F., 2008. Duality-based algorithms for total variation image restoration. UCLA CAM Report 08-33.