DEEP FACTOR MODEL: A NOVEL APPROACH FOR MOTION COMPENSATED MULTI-DIMENSIONAL MRI

Yan Chen[*], **James H. Holmes**[*], **Curtis Corum**[‡], **Vincent Magnotta**[*], **Mathews Jacob**[*]

[*]University of Iowa.

[‡]Champaign Imaging, LLC

Abstract

Recent quantitative parameter mapping methods including MR fingerprinting (MRF) collect a time series of images that capture the evolution of magnetization. The focus of this work is to introduce a novel approach termed as Deep Factor Model(DFM), which offers an efficient representation of the multi-contrast image time series. The higher efficiency of the representation enables the acquisition of the images in a highly undersampled fashion, which translates to reduced scan time in 3D high-resolution multi-contrast applications. The approach integrates motion estimation and compensation, making the approach robust to subject motion during the scan.

Index Terms—

Multi-Contrast; Motion Correction

1. INTRODUCTION

The signal in MRI is sensitive to multiple physical properties of the tissue, which is exploited to observe organs with different contrast mechanisms. The classical MRI methods approach is to carefully choose the acquisition parameters to make the images sensitive to a specific physical property while being less sensitive to others. For example, the magnetization prepared rapid gradient echo (MPRAGE)[1] sequence uses inversion pulses and restricts the acquisition of the data at specific inversion times to obtain good contrast between the gray matter (GM) and white matter (WM) regions. However, a challenge with this approach is its low acquisition efficiency, resulting from the long waiting times and delay times to allow the magnetization to recover. The MPnRAGE[2] sequence was introduced recently, which relies on continuously acquiring the images using radial acquisitions after the inversion pulse; the radial data is binned depending on the inversion time and recovered using gridding. In addition to improving acquisition efficiency, this approach enables the acquisition of images at multiple inversion times.

The MPnRAGE approach has conceptual similarities to MRF [3] that continuously acquires the data using incoherent sampling patterns, while magnetization is continuously evolving. MRF attempts to estimate the physical parameters (e.g. T_1, T_2) directly from the undersampled k-t space data using pattern matching. These methods rely on large

dictionaries of MRF derived using Bloch equation simulations, which correspond to different values of the physical parameters. Several constrained reconstruction algorithms were introduced to improve the reconstruction of the individual frames in MRF; the strategies include low-rank factor modeling of the reconstructed images [4].

The main focus of this paper is to introduce DFM for the joint recovery of inversion recovery data. The proposed DFM, illustrated in Fig. 1 (b), is a generalization of the low-rank model. It capitalizes on the benefits of convolutional neural networks (CNN) in representing images. While unrolled algorithms [5] were introduced to improve low-rank or subspace models, these methods need multiple copies of the subspace images at each unrolling step; the extremely high memory demand of these approaches makes them infeasible in our 3D setting.

The DFM represents each image in the time series as the output of a conditional CNN. The input of the CNN is a series of representative image volumes corresponding to a few inversion delays obtained from gridding. The inversion time-dependent network derives each image in the time series as a denoised non-linear function of the gridded images. The temporal factors derived from a dense network are used to modulate the features of the CNN. Unlike supervised deep learning models, DFM is an unsupervised learning approach that directly learns the subject-specific representation from the measured k-t space data. This approach of learning the network parameters from the undersampled measurements is motivated by [6], which learns an image or an ensemble of images from undersampled measurements. We hypothesize that non-linear deep representation is more efficient that the linear low-rank factor models, and offers implicit spatial regularization due to the implicit bias of CNN blocks towards images.

We also capitalize on the unsupervised learning strategy to compensate for subject motion. In particular, we model the subject motion during the scan as a rigid body. We use the relation between rigid body motions and corresponding transformations in the Fourier domain to absorb them into the forward model. These time-varying parameters are assumed to be unknowns and are solved during signal recovery.

The data is acquired using a radial ultrashort echo time (UTE) sequence with intermittent inversion pulses and a delay time for magnetization recovery. We acquire the data from 3T scanners, with and without motion.

2. PROPOSED APPROACH

2.1. Signal Acquisition and Modeling

The acquisition of data is illustrated in Fig. 2. We assume that the evolution of magnetization during each inversion block is identical, and one initial inversion and segmented UTE scheme is performed to reduce steady-state effects. We consider the image volume ρ_τ to be dependent on the delay $\tau(t)$ from the previous inversion pulse.

We model the signal measurements at the time instant

$$\mathbf{y}_t = \mathbf{A}_t\left(\rho_{\tau(t)}\right) + \mathbf{n}_t$$

(1)

Here, $\mathbf{A}_t$ denotes the forward model that accounts for multichannel Fourier measurements and the k-space trajectory at time t, which is measured from the start of the acquisition. In the absence of motion, the measurements y_t corresponding to a specific τ value can be pooled together.

2.2. Deep Factor Model

We note that low-rank factor models represent the signal as

$$\rho(\mathbf{r}, \tau) = \mathbf{U}(\mathbf{r})\mathbf{V}(\tau),$$

(2)

where the columns of $\mathbf{U}(\mathbf{r})$ are the spatial factors and the rows of $\mathbf{V}(\tau)$ are the temporal factors. The spatial basis functions are modulated by the temporal basis functions to generate the image at a specific time τ. Note that low-rank factor models do not use any spatial regularization on the spatial basis functions $\mathbf{U}(\mathbf{r})$.

We model the image ρ_τ as the output of a deep conditional network $\mathcal{N}_{\phi,\theta}$

$$\rho(\mathbf{r}, \tau) = \mathcal{N}_{\phi,\theta}(\gamma, \tau)$$

(3)

ϕ and θ denote the parameters of $\mathcal{N}$, while τ denotes the inversion delay. γ is a coarse initial reconstruction.

We use a deep factor architecture to realize the conditional model as shown in Fig. 1 In particular, we feed delay time τ to a dense network $\mathcal{V}_\theta$, which provides the temporal factors

$$\mathbf{v}(\tau) = \mathcal{V}_\theta(\tau).$$

(4)

We apply channel-wise multiplication between the feature maps and corresponding temporal factors to modulate the feature maps of $\mathcal{U}_\phi$ and obtain

$$\rho(\mathbf{r}, \tau) = \mathcal{U}_\phi(\gamma(\mathbf{r}), \mathbf{v}(\tau))$$

(5)

One may view this approach as selectively activating and suppressing specific channels, based on the delay τ. If the network $\mathcal{U}_\phi$ only consists of a single hidden layer, this approach reduces to the traditional low-rank model. We expect the deep factorization at different

layers to offer improved representation power. The unknowns of (3) are ϕ and θ. In the absence of motion, we pose the reconstction as

$$\{\theta*, \phi*\} = \arg \min_{\theta, \phi} \sum_t \left\| \mathbf{y}_t - \mathbf{A}_t(\mathcal{N}_{\theta, \phi}(\gamma, \tau)) \right\|^2$$

(6)

2.3. Deep Factor Model with Motion Compensation

We assume the brain to be a rigid body and model any subject motion using 3 rotations and 3 translations. The object at time instant t is assumed to be a rotated and translated version of $\rho_{\tau(t)}$. We denote the transformation operator as T_{v_t, δ_t}, where v_t and δ_t are the rotation and translation parameters at time t. In particular,

$$T_{v, \delta}(\rho(\mathbf{r})) = \rho(R_v \mathbf{r} + \delta)$$

(7)

where $\mathbf{r}$ is the coordinate in the image domain. With this assumption, (1) changes as

$$\mathbf{y}_t = \mathbf{A}_t\left(T_{v_t, \delta_t}\left(\rho_{\tau(t)}\right)\right) + \mathbf{n}_t$$

(8)

We note that a rotation in the image domain by an angle v_t corresponds to a rotation by the same angle in the Fourier domain. This amounts to rotating the k-space trajectory by the same angle. Similarly, a translation in the image domain corresponds to a phase modulation in the Fourier domain. Using these Fourier properties, we rewrite the above relation as

$$\mathbf{y}_t = \mathbf{A}_{t, v_t, \delta_t}\left(\rho_{\tau(t)}\right) + \mathbf{n}_t$$

(9)

This reformulation allows us to absorb the impact of motion into the forward model and further extend the proposed method to DFM with motion compensation(DFM-MC). In the presence of motion, we use (9) to minimize the cost function

$$C(\theta, \phi, v, \delta\} = \sum_t \left\| \mathbf{y}_t - \mathbf{A}_{t, v_t, \delta_t}(\mathcal{N}_{\theta, \phi}(\gamma, \tau)) \right\|^2 + \lambda_1 \left\| \nabla v_t \right\|^2 + \lambda_2 \left\| \nabla \delta_t \right\|^2$$

(10)

with respect to the network parameters θ, ϕ as well as the motion parameters v, δ.

3. IMPLEMENTATION DETAILS

The data were acquired from two human volunteers on a 3.0T GE Premier scanner with 48 channel head coil. From each subject, we acquired two datasets. During the first acquisition,

the subjects were instructed to stay still during the 4.3-minute acquisition. During the second acquisition, the subject was instructed to move the head multiple times during the scan. We implemented $\mathbf{A}$ using a multichannel NUFFT operator, which was differentiable with respect to the input image as well as the trajectory parameters. Post recovery, we fed the network $\mathcal{N}$ with different delays τ to generate the images. We used principal component analysis (PCA) to reduce the 48 coils to 10 virtual coils.

3.1. Deep Factor Model

We implemented $\mathcal{U}_\phi$ using three 3D convolutional blocks, each with a 3D convolutional layer, a channel-wise multiplication layer, and a non-linear activation layer. Each block has 16 channels, except the last layer which has 2 channels representing the real and imaginary parts of the output image. $\mathcal{V}_\theta$ contains two 2D convolutional blocks with kernel size 1 and 32 channels. The first block of $\mathcal{U}_\phi$ is activated by tanh function. Other blocks are activated by the Leaky ReLU function. ϕ and θ are randomly initialized.

We binned the data into eight subsets according to $\tau(t)$ and gridding was applied to each subset to obtain 8 volumes, denoted by γ. While these approximate reconstructions are very noisy and exhibit significant alias artifacts, they capture the contrasts reasonably well. In this work, we considered the recovery of images with 32 different inversion times; $\tau(t)$ was linearly sampled from 0 to 1 with 32 steps. We jointly optimized ϕ and θ based on Eq. 6 In this section, we assume the acquisition was collected without motion.

3.2. Deep Factor Model with Motion Compensation

We assume the motion after one inversion pulse remains to be the same and hence estimate the motion of the subject with temporal resolution=4s. The parameters δ_t and ν_t were initialized as zeros in (10) and solved with θ and ϕ.

4. RESULTS

4.1. Validation of Deep factor model

We first validate DFM in the absence of motion in Fig. 3. We compare DFM against the low-rank approach with two ranks. We observed that the lower rank translated to a poor fit to the magnetization recovery curve. In particular, we note that GM does not appear fully inverted in $r = 4$ low-rank reconstructions at $\tau = 550 - 660$ ms. By contrast, the DFM and $r = 8$ low-rank reconstructions show nulled GM as expected.

The reconstructed images demonstrate that the DFM can offer less noisy reconstructions, especially the bins corresponding to $\tau = 330 - 440$ ms and $\tau = 550 - 660$ ms, where WM and GM are nulled. We note that low-rank reconstructions exhibit significant noise, while the DFM reconstructions are relatively less noisy. The plots of the voxel intensity profiles, shown in Fig. 3 (b), show that the DFM curves are closer to the expected magnetization recovery. While the accuracy of the low-rank curves would improve with rank, this will come with increased sensitivity to noise.

4.2. Validation of motion compensation

The results in Fig. 4 show the motion-compensated recovery using DFM-MC. The results demonstrate that the proposed DFM-MC can compensate for motion effects and offer reconstructions that are comparable to DFM reconstructions of the dataset without motion.

5. CONCLUSION

We introduced DFM to jointly recover multi-contrast images from inversion recovery MRI data. The proposed approach further extended and generalized the low-rank method and recovered a series of images. Our experiments demonstrate that the improved representation translated to higher-quality reconstructions than low-rank models. In particular, the DFM reconstructions are less noisy and offer a more faithful representation of the magnetization recovery. The reconstructions also show the great potential of modeling and correcting for motion during the acquisition, which would be beneficial while imaging older subjects.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted using human subject data. The institutional review board at the local institution approved the acquisition of the data, and written consent was obtained from all participants.

Acknowledgments

REFERENCES

[1]. Brant-Zawadzki Michael, Gillan Gary D, and Nitz Wolf-gang R, "Mp rage: a three-dimensional, t1-weighted, gradient-echo sequence-initial experience in the brain.," Radiology, vol. 182, no. 3, pp. 769–775, 1992. [PubMed: 1535892]

[2]. Kecskemeti Steven, Samsonov Alexey, Hurley Samuel A, Dean Douglas C, Field Aaron, and Alexander Andrew L, "Mpnrage: A technique to simultaneously acquire hundreds of differently contrasted mprage images with applications to quantitative t1 mapping," Magnetic resonance in medicine, vol. 75, no. 3, pp. 1040–1053, 2016. [PubMed: 25885265]

[3]. Ma Dan, Gulani Vikas, Seiberlich Nicole, Liu Kecheng, Sunshine Jeffrey L, Duerk Jeffrey L, and Griswold Mark A, "Magnetic resonance fingerprinting," Nature, vol. 495, no. 7440, pp. 187–192, 2013. [PubMed: 23486058]

[4]. Zhao Bo, Setsompop Kawin, Adalsteinsson Elfar, Gagoski Borjan, Ye Huihui, Ma Dan, Jiang Yun, Grant P Ellen, Griswold Mark A, and Wald Lawrence L, "Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling," Magnetic resonance in medicine, vol. 79, no. 2, pp. 933–942, 2018. [PubMed: 28411394]

[5]. Sandino Christopher Michael, Ong Frank, Iyer Siddharth Srinivasan, Bush Adam, and Vasanawala Shreyas, "Deep subspace learning for efficient reconstruction of spatiotemporal imaging data," in NeurIPS 2021 Workshop on Deep Learning and Inverse Problems, 2021.

[6]. Zou Qing, Ahmed Abdul Haseeb, Nagpal Prashant, Kruger Stanley, and Jacob Mathews, "Dynamic imaging using a deep generative storm (gen-storm) model," IEEE transactions on medical imaging, vol. 40, no. 11, pp. 3102–3112, 2021. [PubMed: 33720831]
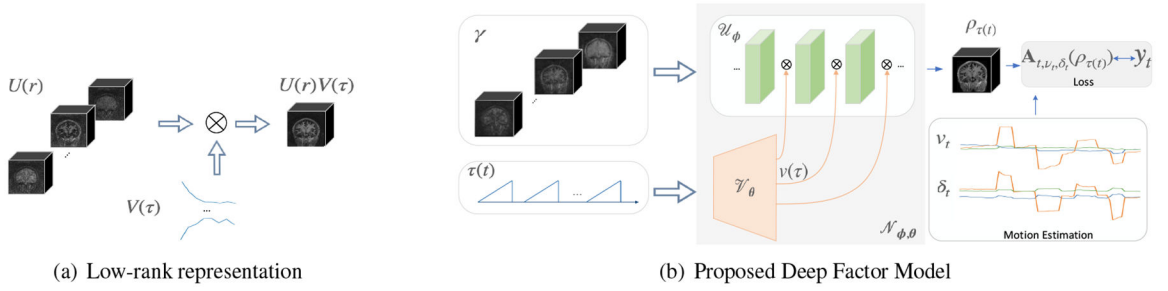
(a) Low-rank representation   (b) Proposed Deep Factor Model

Fig. 1.

Illustration of the low-rank model and DFM. (a) The low-rank method represents the image corresponding to the delay τ as the weighted linear combination of the spatial basis functions $\mathbf{U}(r)$ based on weights $\mathbf{V}(\tau)$; (b) DFM models the signal $\rho(\mathbf{r}, \tau)$ as the output of a conditional network $\mathcal{N}_{\phi,\theta}$. The features of the CNN $\mathcal{U}_\phi$ are modulated by $\upsilon(\tau)$ generated by a dense network $\mathcal{V}_\theta$. The input to $\mathcal{V}_\theta$ is τ, while that of the $\mathcal{U}_\phi$ is an approximate initialization γ obtained from gridding.
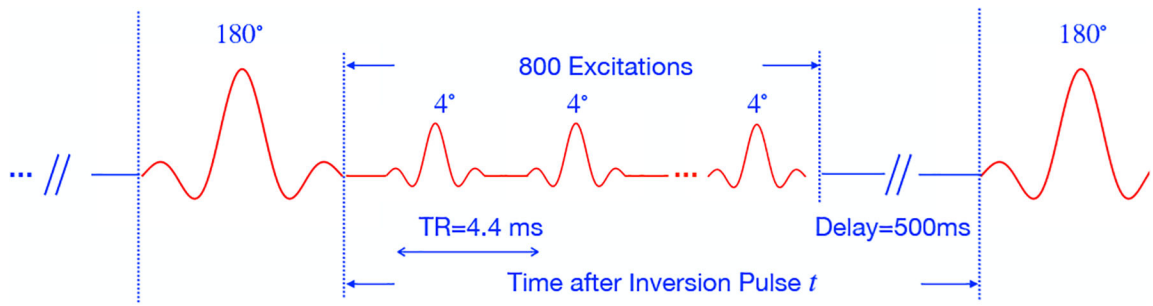
Fig. 2.
Illustration of the pulse sequence. The data is acquired using an inversion recovery with a segmented readout consisting of a train of UTE 3D radial acquisitions with $TR = 4.4$ ms. Intermittent adiabatic inversion pulses are applied after 800 UTE radial lines, with a delay of 500 ms after each segmented acquisition block to allow for signal recovery to a steady state. The total acquisition time was 4.3 minutes to acquire 50K spokes with a matrix size of 256^3. The radial spokes are ordered according to the tiny golden angle view order. The field of view was 24cm^3.
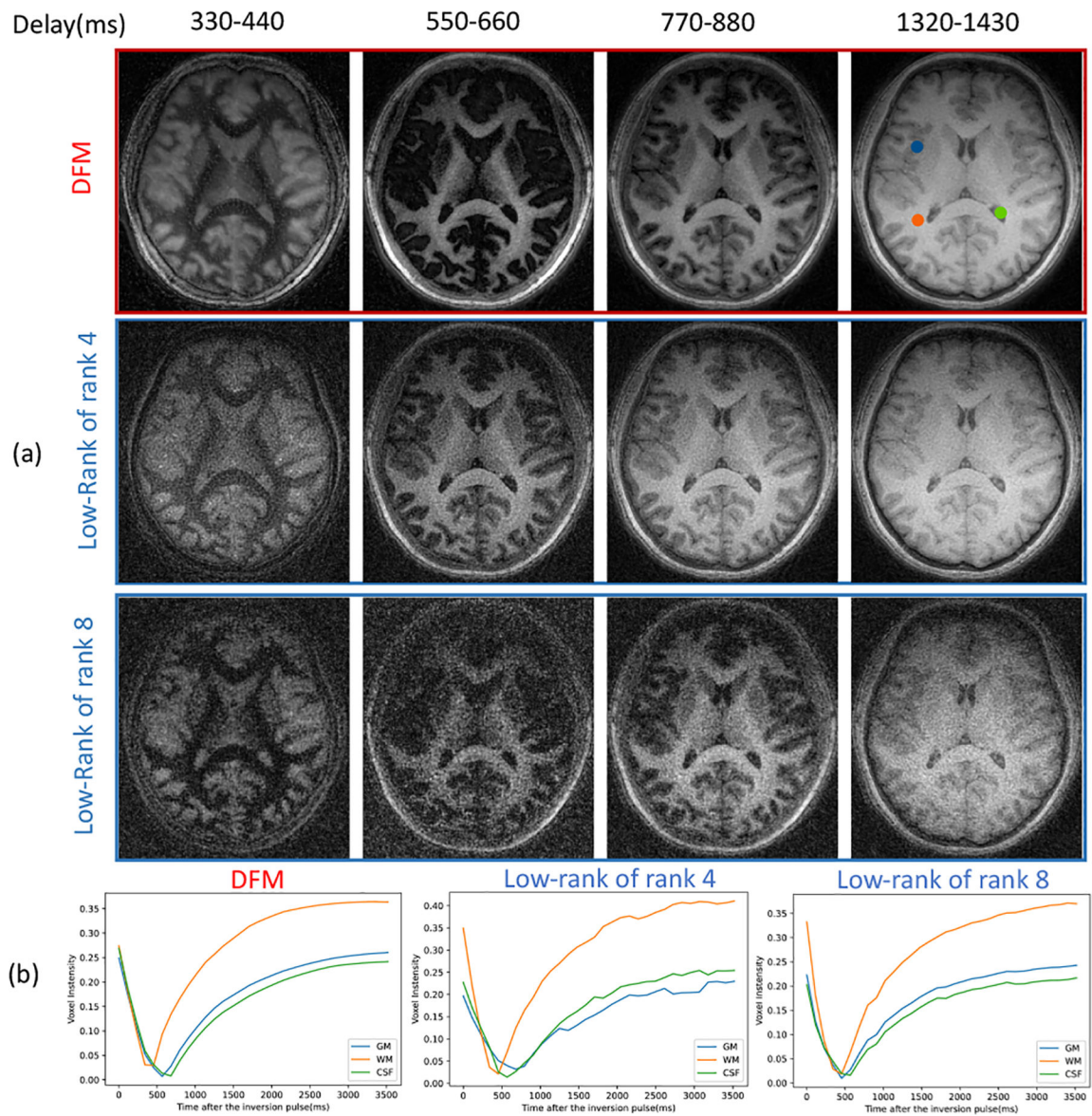
Fig. 3.

Validation of DFM in the absence of motion. (a) The four columns correspond to four delay times τ after one inversion pulse; (b) Magnetization recovery curve of GM, WM, and CSF voxels indicated in the top image, estimated by different approaches.
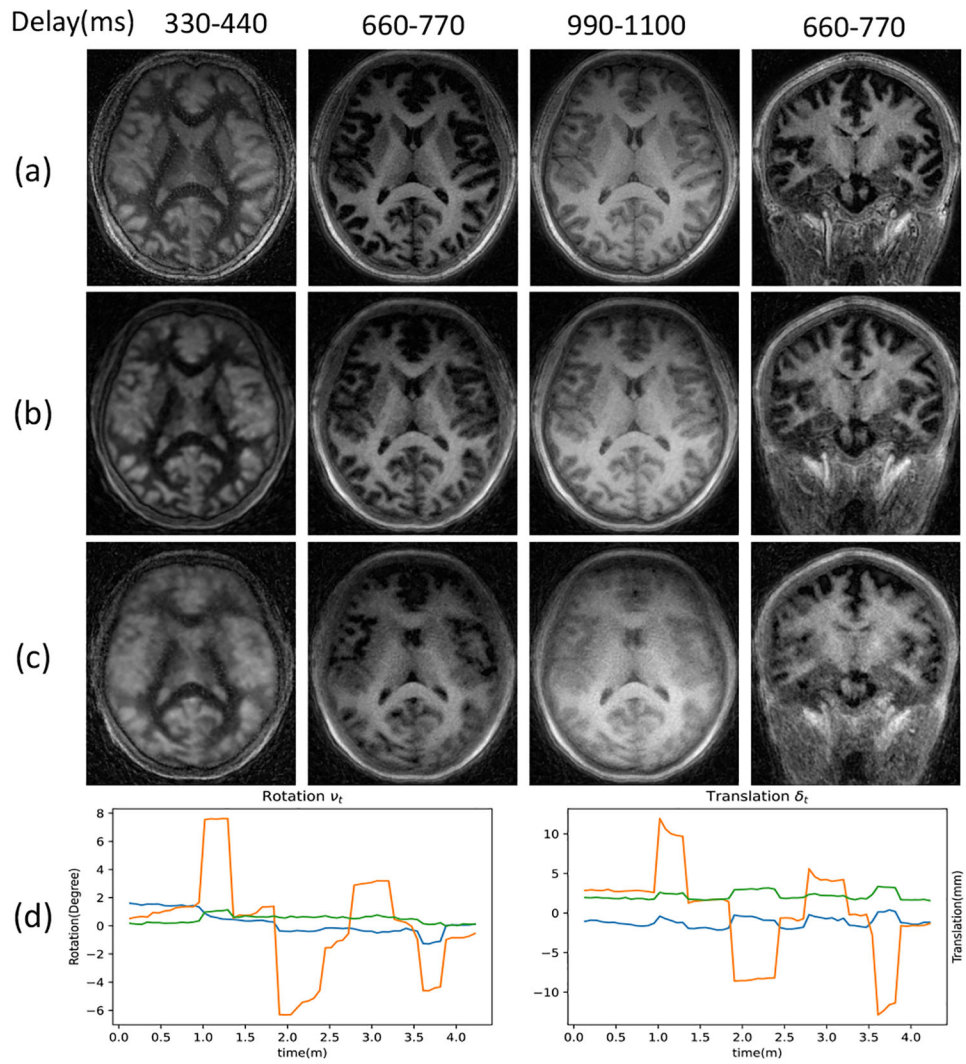
Fig. 4.
Validation of DFM-MC. (a)(c) DFM reconstructions of the acquisitions with or without
motion, respectively; (b) DFM-MC reconstructions of the acquisition with motion; (d)
Motion parameters estimated by DFM-MC including rotation and translation to correct (b)
from (c).