Multi-resolution 3D Nonrigid Registration via Optimal Mass Transport on the GPU

Tauseef ur Rehman[1], Gallagher Pryor[2], John Melonakos[1], and Allen Tannenbaum[1]

[1] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA tauseef,jmelonak,tanneba@ece.gatech.edu,
[2] College of Computing, Georgia Institute of Technology, Atlanta, GA donovang@cc.gatech.edu

Abstract. In this paper we present computationally efficient implementation of the minimizing flow approach for optimal mass transport (OMT) with applications to non-rigid 3D image registration. Our implementation solves the OMT problem via multi-resolution, multigrid, and parallel methodologies on a consumer graphics processing unit (GPU). Although computing the optimal map has shown to be computationally expensive in the past, we show that our approach is almost two orders magnitude faster than previous work and is capable of finding transport maps with optimality measures (mean curl) previously unattainable by other works (which directly influences the accuracy of registration). We give results where the algorithm was used to compute non-rigid registrations of 3D synthetic data as well as intra-patient pre-operative and post-operative 3D brain MRI datasets.

1 Introduction

Image registration and morphing are amongst the most common image processing problems. Registration is the process of establishing a common geometric reference frame between two or more image data sets and is necessary in order to compare or integrate image data obtained from different measurements. A vast amount of literature exists on image registration techniques and we refer the reader to [1, 2] for an overview of this field. In this paper, we approach the registration task by treating it as an optimal mass transport problem. As with other registration techniques, the computational burden associated with this problem is high. We propose a multi-resolution approach for the solution of this problem on the GPU to alleviate this difficulty.

The optimal mass transport problem was first formulated by a French mathematician Gasper Monge in 1781, and was given a modern formulation in the work of Kantorovich [3] and, therefore, is now known as the Monge-Kantorovich problem. The original problem concerned finding the optimal way to move a pile of soil from one site to another in the sense of minimal transportation cost. Hence, the Kantorovich-Wasserstein distance is also commonly referred to as the Earth Mover's Distance (EMD).

Recently, Haker *et al.* [4, 5] have applied the optimal mass transport approach to certain medical image registration problems. Rigorous mathematical details for their algorithm are given by Angenent *et al.* [6]. Although there have been a number of algorithms in the literature for computing an optimal mass transport, the method by Haker *et al.* computes the optimal warp from a first order partial differential equation, which is a computational improvement over earlier proposed higher order methods and computationally complex discrete methods based on linear programming. However, at large grid sizes and especially for 3D registration the computational cost of even this method is significant.

Though computationally expensive, the OMT method has a number of distinguishing characteristics: **(1)** it is a parameter free method and no landmarks need be specified, **(2)** it is symmetrical (the mapping from image A to image B is the inverse of the mapping from B to A), **(3)** its solution is unique (no local minima), **(4)** it can register images where brightness constancy is an invalid assumption, and **(5)** OMT is specifically designed to take into account changes in densities that result from changes in area or volume.

Contribution. In this paper we extend our previous work [7] and implement the more general formulation of the OMT problem for 3D non-rigid registration based on multi-resolution techniques and using the parallel architecture of the GPU. Although multi-resolution methods have served as critical pieces of registration algorithms in the past, it had yet to be shown that the Optimal Mass Transport problem could be solved in the same manner. Our experimental results show that this is indeed the case, a result which has implications for many fields beyond imaging due to the ubiquitous nature of the OMT problem. We also show that the PDE-based solution to the OMT problem is greatly enhanced by our approach to such an extent that it becomes practical for use on large 3D datasets both in terms of speed and accuracy. Overall, these results show that OMT-based image registration is practical on medical imagery and, thus, merits further investigation as an elastic registration technique without the need of smoothness priors or brightness constancy assumptions.

2 Optimal Mass Transport for Registration

2.1 Formulation of the Problem

We will briefly provide an introduction to the modern formulation of the Monge-Kantorovich problem. We assume we are given, a priori, two sub-domains Ω_0 and Ω_1 of R^d with smooth boundaries, and a pair of positive density functions, μ_0 and μ_1 defined on Ω_0 and Ω_1 respectively. We assume that,

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1 \tag{1}$$

This ensures that we have same total mass in both the domains. The functions μ_0 and μ_1 in this formulation can be the same as the source and target images, respectively, or a smooth version of them. They can also be scalar fields that are

appropriate for the underlying physical model. We now consider *diffeomorphisms* $\tilde{u}$ from Ω_0 to Ω_1 which map one density to other in the sense that,

$$\mu_0 = |D\tilde{u}|\mu_1 \circ \tilde{u} \tag{2}$$

which we call the mass preservation (MP) property, and write $\tilde{u} \in MP$. Equation (2) is called the *Jacobian equation.* Here, $|D\tilde{u}|$ denotes the determinant of the Jacobian map $D\tilde{u}$, and $\circ$ denotes composition of functions. It basically implies that if a small region in Ω_0 is mapped to a larger region in Ω_1, then there must be a corresponding decrease in density in order for the mass to be preserved. There may be many such mappings, and we want to pick an optimal one in some sense. Accordingly, we define the squared L^2 Monge-Kantorovich distance as following:

$$d_2^2(\mu_0, \mu_1) = inf_{\tilde{u} \in MP} \int_{\Omega_0} \parallel \tilde{u}(x) - x \parallel^2 \mu_0(x) dx \tag{3}$$

The *optimal MP map* is a map which minimizes this integral while satisfying the constraint given by Equation (2). The Monge-Kantorovich functional, Equation (3), is seen to place a penalty on the distance the map $\tilde{u}$ moves each bit of material, weighted by the material's mass. A fundamental theoretical result [8, 9], is that there is a unique optimal $\tilde{u} \in MP$ transporting μ_0 to μ_1, and that $\tilde{u}$ is characterized as the gradient of a convex function ω, i.e., $\tilde{u} = \nabla\omega$. This theory translates into a practical advantage, since it means that there are no non-global minima to stall our solution process.

2.2 Computing the Transport Map

We will describe here only the algorithm for finding the optimal mapping $\tilde{u}$. The details of this method can be found in [4]. The basic idea for finding the optimal warping function is first to find an initial MP mapping u^0 and update it iteratively to decrease an energy functional. When the pseudo time t goes to ∞, the optimal u will be found, which is $\tilde{u}$. Basically there are two steps. The first step in this algorithm is to find an initial mass preserving mapping. This can be done for general domains using the method of Moser [10] or the algorithm proposed in [4]. The later method can simply be interpreted as the solution of a one-dimensional Monge-Kantorovich problem in the x-direction followed by the solution of a family of one-dimensional Monge-Kantorovich problems in y-direction and finally solve a family of 2D Monge-Kantorovich problems in the z-direction. The second step is to adjust the initial mapping found above iteratively using gradient descent in order to minimize the functional defined in Equation (3), while constraining u so that it continues to satisfy Equation (2). This process iteratively removes the curl from the initial mapping u and, thereby, finds the polar factorization of u. For details on this technique, please refer to [4]. The overall algorithm is summarized graphically in Figure 1. This same algorithm can be used to compute transport map in arbitrary dimensions the only difference being that in R^2 the problem is a bit simpler where you solve

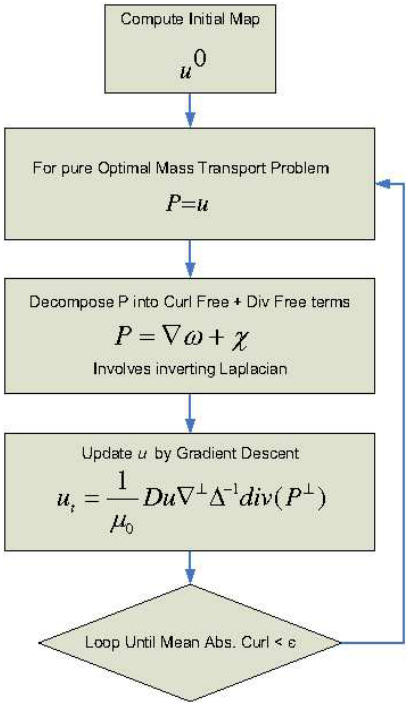

Fig. 1. Optimal Mass Transport Algorithm

the Laplace equation with Dirichlet boundary conditions as compared to solving a Poisson equation with Neumann boundary conditions in higher dimensions. These computations are done in our implementation using Multigrid methods.

3 Implementation

3.1 Multi-resolution Warping

Performing image registration using a multi-resolution approach is widely used to improve speed, accuracy, and robustness. The basic idea is that registration is first performed at a coarse scale. The spatial mapping determined at the coarse level is then used to initialize registration at the next finer level. This process is repeated until it reaches the finest scale. This *coarse-to-fine* strategy greatly improves the registration success rate and also increases robustness by eliminating local optima at coarse scales [11]. Our coarse to fine hierarchy is comprised of three levels (Figure 2).

In our experiments, we found that the coarse-to-fine strategy converges at least twice as fast as the single-resolution solution. Additionally, we found that the coarse to fine method converges to solutions with accuracy (low error metric: mean curl) unattainable by single-resolution methods.

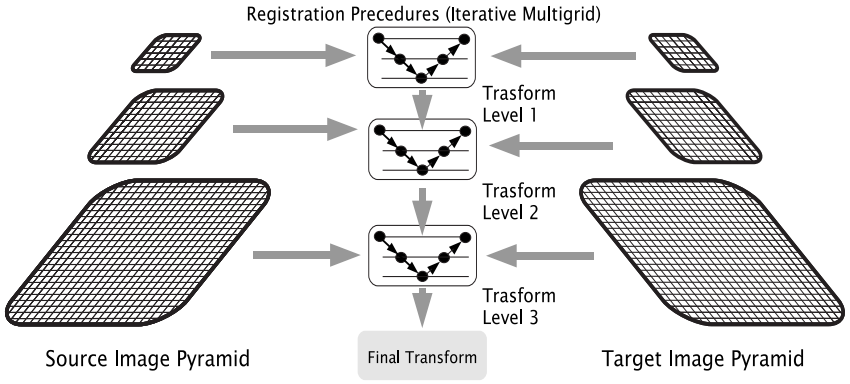

Fig. 2. A Multi-Resolution Registration Scheme. We employ a coarse to fine hierarchy three levels deep with which we solve for an optimal mapping from source to destination data. This method has shown to speed convergence and realize more accurate solutions.

3.2 3D Multigrid Laplacian Inversion

We inverted the Laplacian (a key component of the OMT algorithm) using a 3D multigrid solver. The multigrid idea is very fundamental, it takes advantage of the smoothing properties of the classical iteration methods at high frequencies (Jacobi, Gauss Siedel, SOR etc) and the error smoothing at low frequencies by restriction to coarse grids. The essential multigrid principle is to approximate the smooth (low frequency) part of the error on coarser grids. The non-smooth or rough part is reduced with a small number of iterations with a basic iterative method on the fine grid.

The basic components of multigrid algorithm are discretization, intergrid transfer operators (interpolation & restriction), relaxation scheme and the iterative cycling structure. We used an explicit finite difference scheme for approximating the 3D Poisson equation. This approach uses a 19-point formula on the uniform cubic grid. Relaxation was performed using a parallelizable four-color Gauss-Seidel relaxation scheme. This increases robustness and efficiency and is especially suited for the implementation on the GPU. We used tri-linear interpolation operator for transferring coarse grid correction to fine grids. The residual restriction operator for projecting residual from the fine to coarse grids is the full-weighting scheme. A multigrid *V(2,2)*-cycle algorithm was used to iterate for solution (Residual max norm $\approx 10^{-5}$). Interested readers are referred to [12–14] for details on implementation of the multigrid methods.

3.3 GPU Implementation

An advantage of our solution to the OMT problem is that it is particularly suited for implementation on parallel computing architectures. Over the past few years, it has been shown that graphics processing units (GPUs; now standard in most

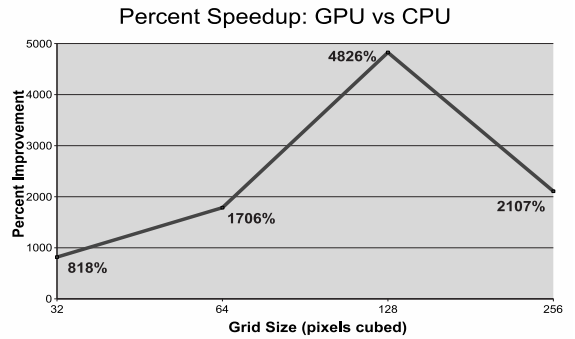

Fig. 3. The GPU realizes an increasing advantage in solving the OMT problem over the CPU as grid size increases up to 128^3 sized grids. Past this point, there is still a large advantage, but a sharp drop is due to memory bandwidth limitations.

consumer-level computers), which are naturally massively parallel, are well suited for these types of parallelizable problems [15, 16].

Taking advantage of these two facts, we implemented our OMT multigrid algorithm on the GPU. The GPU can be considered a massively parallel co-processor and dedicated memory interfacing to the CPU over a standard bus. Modern GPUs are comprised of up to 128 symmetric processors running up to speeds of 1.35Ghz. Their advantage over the CPU in this sense is that while the CPU can execute only one or two threads of computation at a time, the GPU can execute over two orders of magnitude more. Thus, instead of sequentially computing updates on data grids one element at a time, the GPU computes updates on entire grids on each render pass, significantly improving performance (Figure 3). For instance, on a modest Dual Xeon 1.6Ghz machine with an nVidia GeForce 8800 GX GPU (3DMark score of 7200), improvements in speed over our CPU OMT implementation reached 4826 percent on a 128^3 volume data. Presently available GPUs only allow single precision computations, however, this did not affect the stability of the OMT algorithm.

4 Results

We illustrate our registration method using two examples. In the first case, we register a synthetically generated 3D sphere to a deformed (dented) counterpart (Figure 4). In the second case, two 3D brain MRI datasets were registered. The first data set was pre-operative and while the second data set was acquired during surgery and craniotomy and opening of the dura (Figure 5,6). Both were resampled to 256^3 voxels and preprocessed to remove the skull.

In both cases, mean curl of the transport map was reduced to approximately 10^{-3} indicating convergence. However, our coarse-to-fine multigrid implementation on the GPU solves for the optimal transport maps in practical computation times. For instance, registration of the first data (size 128^3) set required 800

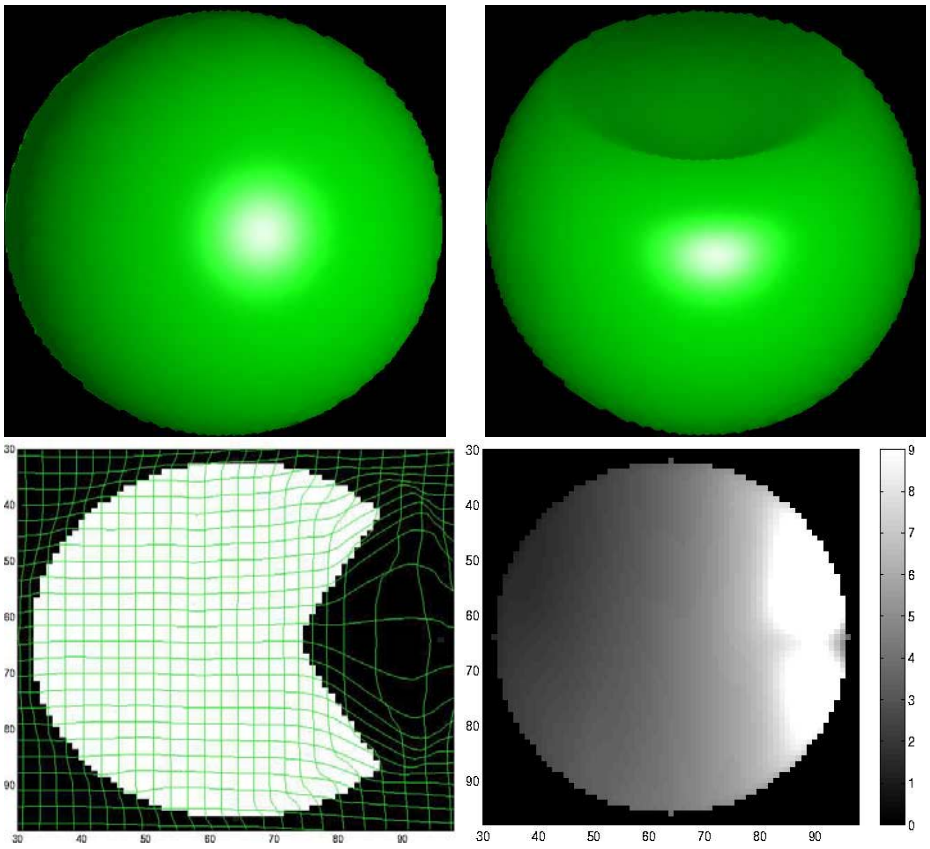

Fig. 4. Synthetic Imagery Results. A sphere is mapped to its deformed counterpart. In the bottom row, the left figure shows the optimal mass transport mapping as a deformation grid overlaid on the destination data. And the right figure shows the magnitude of deformation. Data size 128^3)

iterations of the solver (most at the coarsest scale) requiring less than a minute of computation time. In the second case, (size 256^3) 3600 iterations of the solver were run, requiring less than 15 minutes of computation time.

5 Conclusions

In this paper, we presented a computationally efficient method for 3D image registration based on the classical problem of optimal mass transportation. Many times, global registration methods similar to that presented here are computation intensive making them impractical. However, we have shown that the optimal mass transport is, in fact, a viable solution for elastic registration by achieving low run times on commonly sized 3D datasets on standard desktop computing platforms.

Acknowledgements

This work was supported in part by grants from NSF, AFOSR, ARO, MURI, MRI-HEL as well as by a grant from NIH (NAC P41 RR-13218) through Brigham and Women's Hospital. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`.

References

1. Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–57
2. Brown, L.G.: A survey of medical image registration. ACM Computing Surveys **24** (1992) 325–376
3. Kantorovich, L.V.: On a problem of monge. Uspekhi Matematicheskikh Nauk. **3** (1948) 225–226
4. S.Haker, Zhu, L., Tannenbaum, A., Angenent, S.: Optimal mass transport for registration and warping. International Journal of Computer Vision **60**(3) (2004) 225–240
5. S.Haker, Tannenbaum, A., Kikinis, R.: Mass preserving mappings and image registration. In: MICCAI. (2001) 120–127
6. Angenent, S., Haker, S., Tannenbaum, A.: Minimizing flows for the monge-kantorovich problem. SIAM Journal of Mathematical Analysis **36** (2003) 61–97
7. Rehman, T., Tannenbaum, A.: Multigrid optimal mass transport for image registration and morphing. In: Proceedings of SPIE Conference on Computational Imaging V. Volume 6498. (2007)
8. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Communications on Pure and Applied Mathematics **64** (1991) 375–417
9. Gangbo, W., McCann, R.: The geometry of optimal transportation. Acta Mathematica **177** (1996) 113–161
10. Moser, J.: On the volume elements on a manifold. Transactions of the American Mathematical Society **120** (1965) 286–294
11. Yoo, T.S.: Insight into Images, Principles and Practice for Segmentation, Registration and Image Analysis. A. K. Peters Ltd. 2004 (2004)
12. Gupta, M.M., Zhang, J.: High accuracy multigrid solution of the 3d convection-diffusion equation. Applied Mathematics and Computation **113** (2000) 249–274
13. W.L.Briggs, Hensen, V., McCormick, S.: A Multigrid Tutorial. SIAM (2000)
14. Trottenberg, U., Oosterelee, C., Schüller, A.: A Multigrid Tutorial. SIAM (2000)
15. Bolz, J., et al.: Sparse matrix solvers on the GPU: Conjugate gradients and multigrid. In: Proceedings of SIGGRAPH. Volume 22. (2003) 917–924
16. Nolan, G., et al.: A multigrid solver for boundary value problems using programmable graphics hardware. In: Proceedings of SIGGRAPH. (2003) 102–111

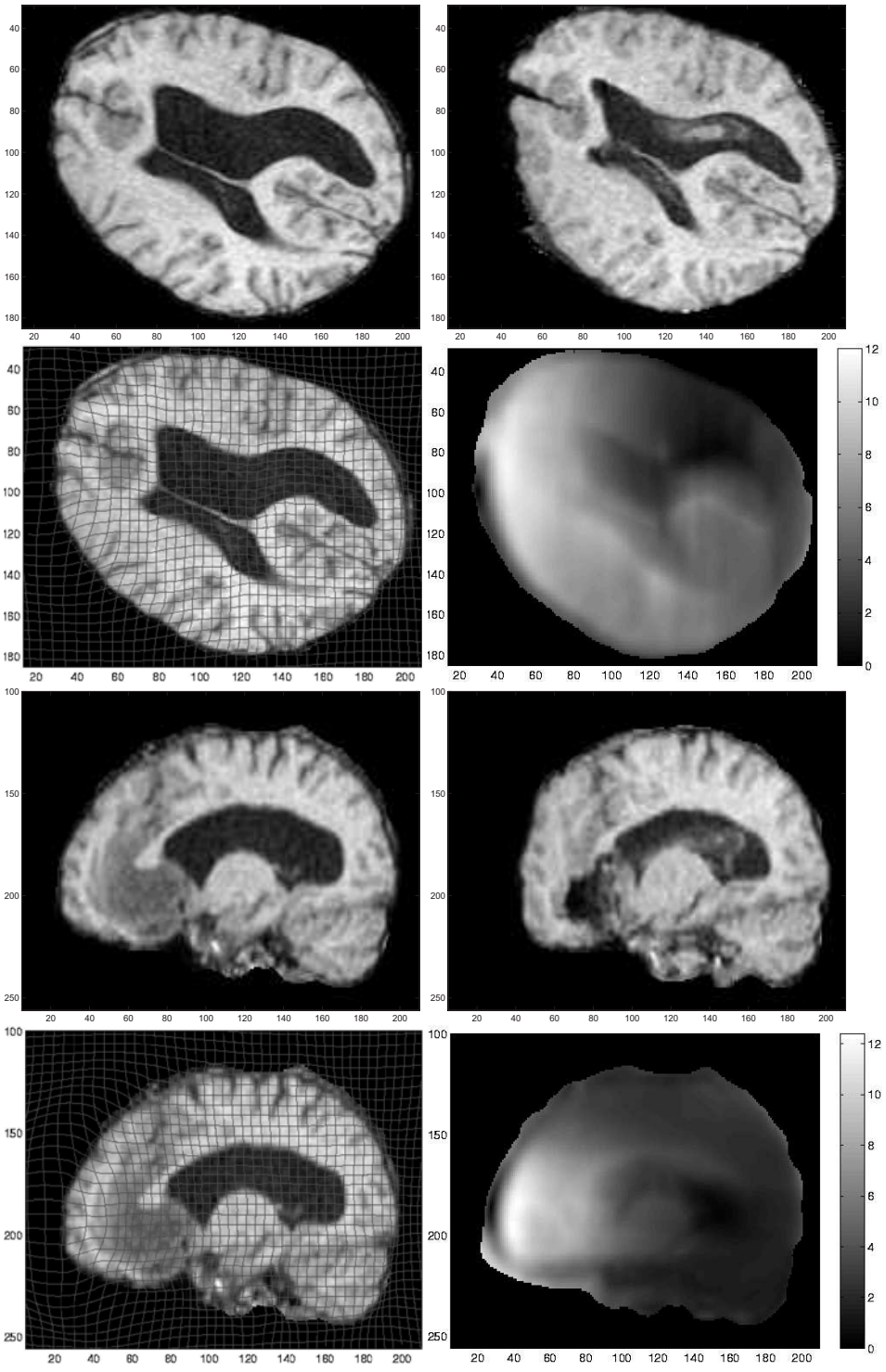

Fig. 5. Brain Sag Registration. The top four figures show the registration results on an axial slice and the bottom four show results for a saggital slice from the 3D volume. The deformation due to the brain sag after carniotomy and openning of the dura is clearly visible in both the deformation grid and the magnitude of deformation plots. The gravity vector is parallel to the horizontal axis. A rigid shift can also be noticed due to slight displacement of the head during surgery.

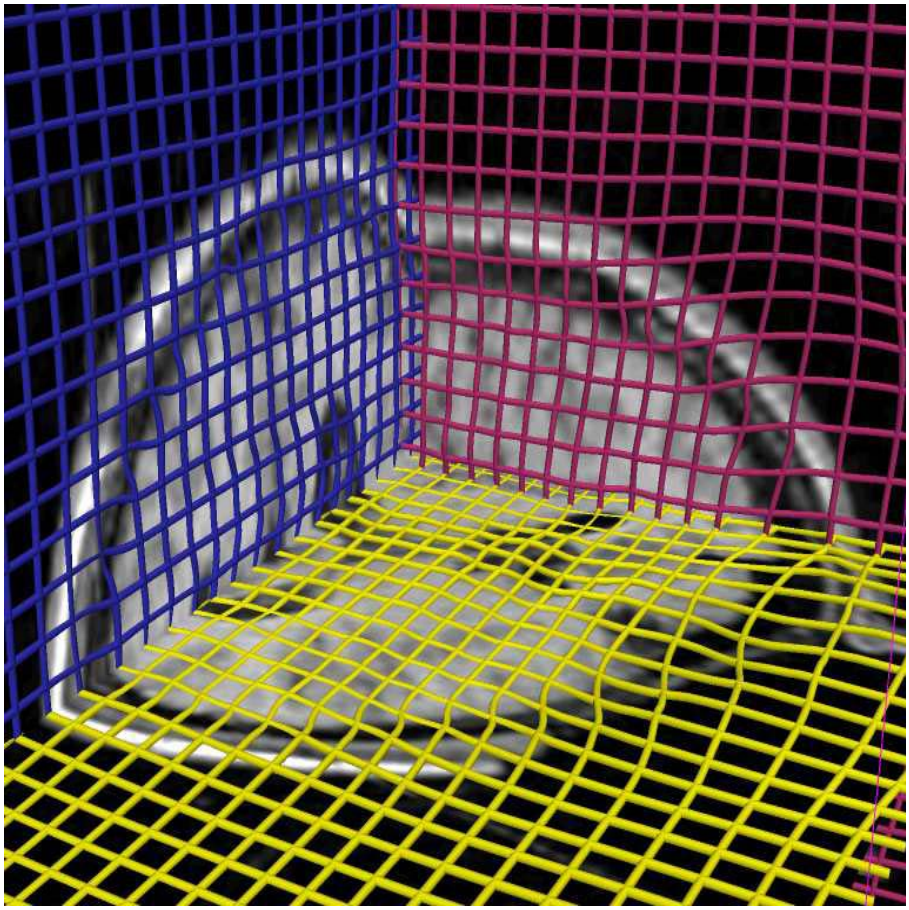

Fig. 6. Brain Sag Registration(3D View). The brain sag is visible in the anterior portion of the brain. (Data size 256^3).