Distributed Global Digital Volume Correlation by Optimal Transport

J. Michael L. MacNeil*, Dmitriy Morozov*, Francesco Panerai†, Dilworth Parkinson‡, Harold Barnard ‡, Daniela Ushizima* §

* Computational Research Division, Lawrence Berkeley National Laboratory
† Department of Aerospace Engineering, University of Illinois Urbana Champaign
‡ Advanced Light Source, Lawrence Berkeley National Laboratory
§ Berkeley Institute for Data Science, University of California Berkeley

***Abstract*—Because of the speed and data rates of time-resolved experiments at facilities such as synchrotron beamlines, automation is critical during time-resolved experiments. In 3D imaging experiments like microCT (μCT), this includes recognizing features of interest and "zooming in" spatially and temporally to those features; ideally without requiring advanced information about which features are being imaged. Digital Volume Correlation (DVC) can achieve this by measuring the deformation field between images, but has not been used during autonomous experiments because of the scalability of the codes. In this work, we propose a model for global DVC and a parallel algorithm for solving it for large-scale images, suitable for giving feedback for autonomous experiments at synchrotron-based microCT beamlines. In particular, we leverage recent advancements in entropy-regularized optimal transport to develop efficient, simple-to-implement, parallel algorithms which scale linearly ($\mathcal{O}(N)$) in space and time, where N is the number of voxels, and well with an increasing number of processors. As a demonstration, we compute the deformation field for every voxel from a μCT volume with dimensions 2560x2560x2160. We discuss implementation details, drawbacks and future directions.**

I. Introduction

MicroCT (μCT) data collection has been used to study carbon fibers [1], [2], ceramic matrix composites [3], [4], batteries, bones, and many other materials [5], in many cases while subjecting materials to conditions of interest, such as high temperature, pressure, and mechanical loads. But there is a growing gap between the rate at which one can acquire volumetric μCT data, which is on the order of seconds to minutes, and the efficiency of algorithms for image analysis. Planned instrument upgrades will lead facilities to produce terabytes of image-based data per second, further enlarging this gap. This is especially problematic because increasingly, it is desirable to run time-resolved experiments autonomously, adjusting the spatial and temporal sampling rate as well as the location being imaged to collect data on the most relevant and interesting parts of the sample which show changes during the experiment.

An algorithm that could span the wide range of samples that are studied and be a key part of autonomous experiments is one that would determine the displacement field [6]. In studying μCT images, this problem is generally referred to as *Digital Volume Correlation* (DVC). Although the method has applications to several other areas, DVC is highly relevant to mechanical testing of materials. When mechanical loads are applied to a material (tension, compression, torsion, etc), the resulting deformation field provides a direct, spatially resolved measurement of important features that contribute to structural behavior, such as failure localization, areas of stress concentration, and bulk properties under load [6].

In current scientific workflows, DVC is used off-line, after an experiment has taken place. The main reason is computational complexity. Determining the displacement field is a challenging optimization problem, and even simplified iterative algorithms are non-trivial to implement for datasets produced by synchrotron beamlines.

The main contribution of this paper is an optimal transport model of global DVC and an efficient, parallel algorithm for solving it. The key observation is that global DVC is equivalent to an optimal transport image registration problem [7], and recently proposed approximate solvers [8] can scale to full-resolution datasets. This work is a step in the direction of generalized autonomous experiments for μCT that can "zoom in" spatially and temporally on regions of interest during an experiment.

The paper is organized as follows. Section II reviews the DVC problem and discusses local and global computational approaches. Section III provides background on Optimal Transport and its application to DVC. Section IV provides background on Entropy-Regularized Optimal Transport and describes algorithms for solving it. Section V describes implementation details. Section VI describes the dataset we use and its acquisition. Section VII presents our results on experiments involving synthetic and real data. We conclude and discuss drawbacks and future directions in Section VIII.

II. Problem Formulation and Related Work

We consider a mass of material captured by an image, called the target image, and denote this as $f(x)$, where x is a position in $\mathbb{R}^3$. A second image (called the reference image) contains the same mass of material, however the material has now been deformed, and the density is distributed according to some new function $g(x)$. Digital Volume Correlation algorithms compute a vector field $\phi : \mathbb{R}^3 \to \mathbb{R}^3$, called the displacement

field, which deforms the reference image into the target image. Mathematically, it is the solution to the following optimization problem [6],

$$\min S(f, g(\phi)) + R(\phi) \tag{1}$$

where S is a measure of similarity between the target image $f(x)$ and the transformed image, $g(\phi(x))$, and $R(\phi)$ is a regularization term. There are many choices for similarity measures and regularizations; see [9], [6], [10], and an extensive review is available in Buljac et al [6].

Two approaches have emerged within the DVC community: local, which were originally proposed, and global, which are more recent refinements. Each approach and the opportunities for parallelism are outlined in the next subsections.

A. Local Approaches

Local approaches describe a set of algorithms to DVC which approximately solve Equation (1), processing local subvolumes independently. The main idea is as follows. A subvolume in the reference image, centered at a point x_i with window size w and volume w^3, is extracted. For this subvolume, the best fitting affine transformation to the target subvolume, in some neighbourhood of x_i, is computed. The resulting affine transformation is used for the deformation $\phi(x)$ at x_i.

This approach is particularly amenable to parallelism as each subvolume can be handled independently. For this reason, to the best of our knowledge, this is the only approach which has been used for DVC on HPC systems [11], [12]. However, local approaches do not account for assignments of neighbouring subvolumes, and this can lead to ambiguous assignments. This shortcoming, and several other drawbacks, have been documented elsewhere, and we refer the reader in particular to [13], [6].

B. Global Approaches

Global approaches to DVC, and the subject of this work, look to find a continuous $\phi : \mathbb{R}^3 \to \mathbb{R}^3$, which approximately solves Equation (1) over the entire image volume [14]. This is closely related to image registration algorithms from image processing and applied mathematics, and more faithfully constructs the deformation field [6]. However, global approaches are difficult to scale. A general approach is as follows. The vector field is expanded in a set of basis functions, which transforms Equation (1) into a nonlinear least-squares problem. Note this problem is not necessarily convex, and so only local minima can be expected. The nonlinear system is then handled using a Gauss-Newton solver [13], which can give a high-order accuracy. These nonlinear least squares solvers, in comparison to the local DVC approach, are non-trivial to parallelize [15].

III. Optimal Transport Model

The model we adopt for DVC is derived from optimal transport, which is a particularly rich mathematical area and whose application to image registration has been explored before [7], [16]. Considering each voxel as having a particular amount of mass, let $P_{i,j}$ be a matrix which dictates the amount of mass from voxel x_i from the target image $f(x)$ to be distributed to y_j in the reference image $g(y)$. This table is called the mass transportation plan [8], [7]. The cost of moving some mass is set as the amount of mass moved, multiplied by the distance squared that it travelled, or $C_{i,j} = P_{i,j}||x_i - y_j||_2^2$. Determining the minimum cost plan that moves all mass from $f(x)$ to $g(y)$ then solves the linear program.

$$\min_{P \in U(f,g)} \sum_{i,j} P_{i,j}||x_i - y_j||_2^2. \tag{2}$$

$$U(f, g) = \{U \in {\mathbb{R}_+}^{N\text{x}N} \mid U\mathbf{1} = f,\ U^T\mathbf{1} = g\}. \tag{3}$$

A deformation map can then be defined by taking the weighted average point [8]

$$\phi(x_i) = \sum_j P_{i,j} y_j / f(x_i). \tag{4}$$

Note in particular this model has been explored in image warping and registration by Haker et al. [7], and enjoys several favorable qualities. There is a unique global minimum, and it is parameter free [7].

The drawback of this approach is that, naively, the memory footprint of $P_{i,j}$ requires $\mathcal{O}(N^2)$ entries, where N is the number of voxels. For volumetric μCT images, with billions of voxels, this would require a large amount of memory, in the exabyte range. In practice, the resulting transportation plan is sparse, and requires much less memory than this, however determining the support *a-priori* is a challenge [17]. Haker et al. [7] cite this as the reason for avoiding the linear programming solution in the first place, motivating a partial differential equations approach. However, recent algorithms are available to approximate (2), which are substantially simpler to implement and parallelize [8]. We provide details in the next section.

IV. Algorithm Description

While there are several efficient approaches to solving the optimal transport problem (see [18] for a recent review), we focus on the entropy-regularized optimal transport proposed by Cuturi [19] because of the opportunities for parallelism. Other possibilities which provide excellent efficiency include the Auction Algorithm [20], Dual Descent [21], a multiscale approach [22], as well as recent results in linear programming [23]. However, as shown in the next section, the entropic regularization approach admits an efficient algorithm which is simple to parallelize. This approach has recently been used in the medical imaging community for diffeomorphic registration [16] and in several other fields including computer graphics and machine learning [24], [17]. It describes an algorithm for approximating the transport plan $P_{i,j}$ in $\mathcal{O}(N)$ space and $\mathcal{O}(N \log N)$ operations. The operations are improved to $\mathcal{O}(N)$ in the implementation section.

A. Entropic Regularization

The main idea behind the method is to find approximate solutions to the linear optimal transport program by regularizing with an entropy term $H(P)$ defined as

$$H(P) = -\sum_{i,j} P_{i,j}\big(\log(P_{i,j}) - 1\big) \tag{5}$$

Adding this to our linear program, the approximation, which we denote P_σ, satisfies

$$P_\sigma^* = \min_{P \in U(f,g)} \langle P, D \rangle - \sigma^2 H(P). \tag{6}$$

Note this regularization now enforces the positivity of P, and so the positivity constraint can be relaxed.

While this problem initially looks more challenging there is a particularly nice form of the solution. One can show that P_σ decomposes into the following product [19]

$$P_\sigma = \mathbf{diag}(u) K_\sigma \mathbf{diag}(v), \tag{7}$$

where

- u and v are called *scaling vectors* and reside in $\mathbb{R}^n$;
- $K_\sigma = \exp(-||x_i - y_j||_2^2/\sigma^2)$.

We note that in this formulation matrix K_σ has a closed form, and we don't need to store it explicitly. Instead we require space for the scaling vectors u and v, reducing memory consumption to $\mathcal{O}(2N)$ from $\mathcal{O}(N^2)$ in the original linear program. This is a dramatic savings in the case of images.

B. Sinkhorn's Algorithm

It still remains to determine the unknown scaling vectors u and v to find the approximate optimal transport plan. Following Cuturi [19], we substitute the expression for P_σ from (7) into the linear constraints in (3) to get

$$u \odot K_\sigma v = f, \quad v \odot K_\sigma u = g. \tag{8}$$

Here, $\odot$ is element-wise multiplication, and $K_\sigma u$ is a matrix-vector product. Moreover, multiplication by K_σ is a Gaussian convolution and can be computed in $\mathcal{O}(N \log N)$ time instead of $\mathcal{O}(N^2)$.

To find a solution (u, v), Sinkhorn's algorithm iterates between updating u and v as follows. Initiate the algorithm with $u^0 = v^0$ and repeat the following until convergence

$$u^k = \rho_o \oslash (K_\sigma v^{k-1}), \quad v^k = \rho_1 \oslash (K_\sigma u^k). \tag{9}$$

Here, $\oslash$ is element-wise division. Therefore, an approximation to the optimal transport plan P_σ can be found by performing Gaussian convolution and element-wise division, each of which can be performed in $\mathcal{O}(N \log N)$ operations.

V. Implementation

In this section we provide the details of our implementation of Sinkhorn's algorithm for large scientific images.

A. Block Parallel Processing

Our distributed implementation of Sinkhorn's algorithm is done using the block parallel data analysis library DIY [25], which follows the bulk synchronous parallel (BSP) model of computation. The main abstraction in DIY is a block, which is an indivisible unit of data, defined by the user. Operations are performed on each block in parallel, interleaved with communication phases between neighbouring blocks.

Blocks for our application are subvolumes with ghost zones set to the half-width of the Gaussian convolution, as detailed in the next subsection. The image is decomposed into the number of subvolumes equal to the number of available processors.

Each iteration of Sinkhorn's algorithm is straightforward to implement using DIY. Gaussian convolution is performed over each block in parallel. Afterwards, the needed overlap from neighbouring block ghost zones is exchanged. The element-wise division is then done in parallel with a similar neighbour exchange, and the process is repeated until convergence.

B. Gaussian Convolution

Operating on blocks we have chosen makes performing Gaussian Convolution using a distributed Discrete Cosine Transform (DCT) challenging. This poses an issue for large deformations, where we need to perform Gaussian convolution over a wide window. Applying the Gaussian kernel directly, with a truncated window, leads to an implementation which scales as $\mathcal{O}(w^3 N)$. Even for modestly sized windows (which are proportionate to expected deformation sizes) this can be prohibitively costly.

Fortunately, there are several Gaussian convolution algorithms which are constant in the window size. For a thorough treatment of these we refer the reader to the excellent survey by Greteuer [26]. Of the available options, we chose the extended Box Car filter approximation. This introduces a small error, as it approximates the Gaussian kernel rather than implementing it exactly. However, in our validation results, we did not find this to be significant.

C. Numerical Instability

A known issue with Sinkhorn's algorithm is that for small values of σ, the values of $K_\sigma u$ can become very small, leading to numerical instability and overflow [8], [27]. One solution is to use log domain iterations to offset this [8], however we found this did not make an improvement in stability for our experiments. Chizat et al. propose to use solutions for larger values of σ, which are easier to obtain, as initial solutions for smaller values of σ [28]. We call the largest value of σ, σ_M and the smallest σ_m. The largest scale has the interpretation of the largest expected deformation, while the smallest scale can be interpreted as the finest observable features.

D. Monitoring Convergence

Monitoring convergence of Sinkhorn's algorithm, particularly for images, is not a straightforward process; it depends on the distribution of pixel values as well as initial scaling.

For this work, we assume the image has been quantized to $[0, 255]$ and stored as an unsigned 8-bit integer. We scale the image so that pixel values belong to $[0, 1]$ and add a small number, 10^{-4}, so as to avoid dividing by zero. Error is then measured by the relative residual

$$r = ||u \cdot Kv - f||_1 / ||f||_1 \tag{10}$$

and the program is terminated if this falls below 0.02. In practice, this condition is satisfied when the difference in the produced output is less than 2% of the mean value of the input, which appears to be comparable to previous work [13].

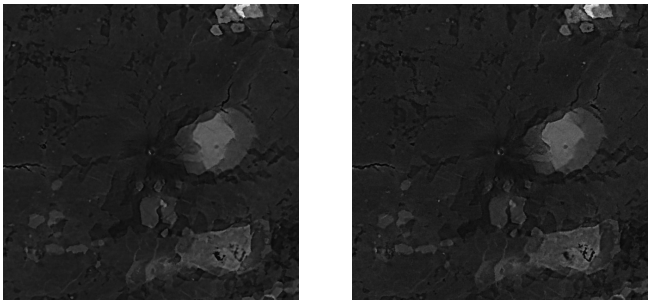

Fig. 1: Example slices of synthetically generated images. Left: Target Synthetic Image. Right: Reference Synthetic Image. Created by using the unloaded and loaded images as scaling vectors in 8. Note each is similar, however fine details in the reference image have been diffused in the target image.

VI. Data Description

The dataset we used consists of two μCT images of L-Chondrite meteorite samples subjected to mechanical compression. The first is an unloaded sample and the second has been compressed by 400μm with a load of $750N$. The dimensions of each image are 2560×2560×x2160 voxels. A more detailed description of the samples and data acquisition is provided below.

The images were collected as part of an in-situ μCT campaign conducted at the beamline 8.3.2 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory. The campaign gathered data in support of the NASA Asteroid Threat Assessment Project (ATAP). As part of ATAP, the internal structure of chondrite meteorites was characterized using μCT and samples were tested under mechanical loads to understand their failure mechanisms. L-type ordinary chondrites are the second most common group of meteorites, and exhibit a low iron content compared to H-chondrites. The experimental setup includes a mechanical testing chamber with a motor driven system to deliver up to 2000 N of compression load. The chamber also features a 1-cm-tall, 0.5-mm-thick axis-symmetric X-ray transmissive window that allows X-ray imaging while applying load to the sample. An L-chondrite sample was subjected to increasing compression loads by applying displacements of 20 μm, alternated by tomography scans under static conditions. Scans were acquired using polychromatic X-rays, 5× magnification Mitutoyo lens, 50 μm LuAG scintillator, and a 2,560×2,160 pixel sCMOS sensor (pco.EDGE), by collecting 1,025 radiographs over 180° rotation, with an exposure time of 400 ms. We obtained scans of 2,560×2,560×2,160 voxels, with a pixel size of 1.25 μm. For the present analysis we tested scaling of our algorithm on the entire 3D volume of the datasets collected at 0 μm (unloaded) and 400 μm displacement (≈750 N).

VII. Results

We present results from two sets of experiments. The first, aimed at verifying the accuracy of Sinkhorns algorithm, is on a synthetic dataset. This set of results was run on a custom workstation, where the entire problem fit into memory. The second are from a large, real dataset, which did not fit into memory. This set of results used a distributed supercomputing system Cori. A description of the hardware is provided in the next subsection, and results presented in the subsequent two.

A. Hardware Platforms

Accuracy tests on the synthetic dataset were performed on a custom workstation with 16 Intel Xeon Gold 5122, 3.60 GHz processors and 96 GB DDR4 2666MHz memory.

Distributed results and performance tests on real data were performed on the Phase 1, "Haswell" partition platform of the Cori system maintained by the National Energy Research Scientific Computing Center (NERSC). Phase 1 of the Cori supercomputer contains 2,388 nodes. Each node contains two-sockets, each socket is populated with a 16-core Intel® Xeon™ Processors E5-2698 v3 ("Haswell") at 2.3 GHz and 128 GB DDR4 2133 MHz memory. With hyper-threading, each node contains 64 logical cores (2 hyper threads per core).

B. Synthetic Data

We evaluated accuracy by comparing with a known, synthetic transportation plan following the work of [29]. This can be done by choosing any scaling vectors and then producing images from the marginal constraints in (8). For example, if we have two images, image one and image two, we can use image one as u and image two as v (8). This produces two *synthetic* images, f and g, on which we can test our algorithm. We note there is some bias in choosing this synthetic test, as this is precisely the form of transportation plan our algorithm aims to recover. However, this also creates phantoms where sharp details in the reference image, have diffused in the second, creating a challenging non-rigid deformation test. We intend to perform substantially more tests on various, known transformations in future work.

To create synthetic images, we take a (640, 640, 580) crop from the center of the loaded and unloaded L-Chondrite images. We then set the scaling vectors, (u, v) in (8) to be these crops. As these transportation maps also depend on σ, we created a transportation plan for sigma values of one through five. An example of the resulting synthetic reference, target image for σ value of 1 is given in Figure 1.

Sinkhorns iteration was performed on the synthetic dataset and the sources of error collected were (a) the final residual, denoted by r and given in (10); (b) the maximum difference between the truth displacement field and the computed displacement field, denoted $||t_d - c_d||_\infty$; (c) the average residual between the truth displacement field and the computed displacement field, denoted $||t_d - c_d||_1$.

A report of these errors is in Table I. Similar to previous work, there are boundary effects which contribute to the maximum error being relatively high [13]. The reconstruction error was generally very low, indicating the transport plan can reproduce either the reference or target image with high accuracy [14], [13].

We also observe good qualitative results to accompany the quantitative ones. For the synthetic images corresponding to $\sigma = 1$, the ground truth deformation map and the computed deformation map is presented in Figure 2. For this value of σ,

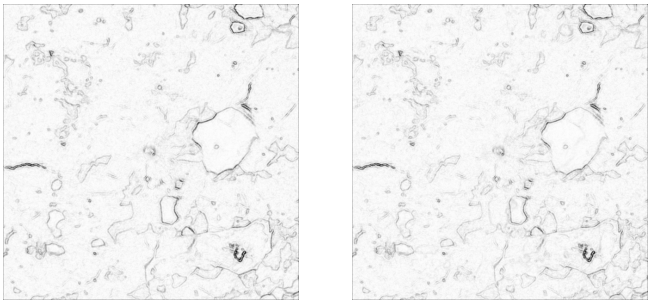

Fig. 2: Example slices showing the magnitude of the deformation map $\phi(x)$ from (4). The ground truth is shown on the left and the result we compute is shown on the right.

the largest value in the ground truth was 1.36. For visualization, the images were scaled by $255/1.36$, clamped at 255 and quantized to 8-bit integers. Qualitatively there is little difference between the true deformation and that which we compute.

C. Real Data

Saved as floating point, each image occupies 55GB of memory, making the memory footprint for the scaling vectors 110GB. We chose the largest scale to be $\sigma_M = 15$ and the lowest to be $\sigma_m = 5$. In each experiment, instability set in at $\sigma = 7$ and the algorithm was terminated.

The final residual for $\sigma = 7$ was still very low, 0.023, indicating an accurate reconstruction of both reference and target images. Figure 4 shows the deformation map rendering.

We note that the current implementation does not subtract off a rigid transformation needed to be able to view finer details such as cracks. Most of the computed deformation seen is overall sample motion. We intend to follow up this work with a more thorough investigation that will allow detection of a region of highest interest based on the analysis of the digital volume correlation results.

D. Performance

Our performance tests examined strong scalability, fixing the size of the dataset and varying the number of processors applied to the problem. A processor, in these results, refers to a logical CPU core. Each MPI rank was assigned a single logical core. The results are presented in Figure 3. We achieve square root scaling up to 8192 processors.

Relative to the time to acquire and reconstruct a microCT image at the beamline, the computation time is acceptable for a number of use cases. The acquisition time of μCT images depends on several aspects of the beamline experiment. A number of experiments, particularly at Beamline 8.3.2 at the

TABLE I: Quantitative Results on Synthetic Data

σ	Residual r	$\|\|t_d - c_d\|\|_1$	$\|\|t_d - c_d\|\|_\infty$
5	0.005	0.047	0.943
4	0.006	0.039	0.897
3	0.007	0.031	0.815
2	0.009	0.023	0.642
1	0.006	0.012	0.525

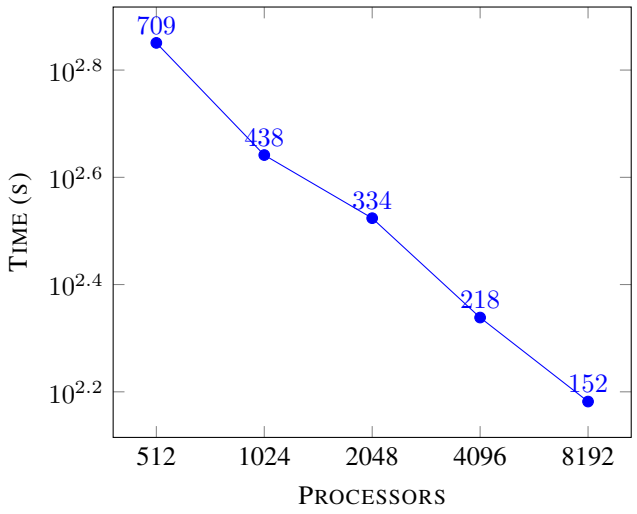

Fig. 3: Strong scaling investigation of DVC. Numbers above the data points indicate time in seconds for convenience.

Advanced Light source, use monochromatic light and require high resolution. Image acquisition, including reconstruction, in this setting can take between 1 and 5 minutes. For this group of experiments our turnaround time is short enough that it fits within the experimental workflow and could help guide the imaging. For lower resolution experiments, acquisition speeds can take as little as 100 milliseconds, which is beyond our current capability and is the subject of future work.

We expect performance to drop off for a larger number of processors due to increased communication from size of the ghost zones for Gaussian blur, particularly with the window sizes considered in our application. As blocks get smaller and the width of the ghost zones is held constant, the information needed for each point extends beyond what is computed by a single neighbour, making the exchange phase more costly.

VIII. Conclusion and Future Directions

We have presented a model for global digital volume correlation which is parallelizable for large scale applications. We have demonstrated that this model can compute challenging non-rigid deformations as well as scale to large scientific datasets produced by μCT. While this current work is already promising, there are several future directions to be pursued. In particular, a more thorough validation of the model with real data for different science applications. In particular, we will need to do further experiments to determine specific experimental feedback loops to be informed by these deformation maps. The ultimate goal of this work is an algorithm which can be deployed at a synchrotron beamline. While we have provided theoretical evidence of connections with global DVC approaches, more empirical evidence validating the model.

IX. Acknowledgements

This work was supported by the Office of Science, of the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231, and by the use of resources of the National Energy Research Scientific Computing Center (NERSC) and

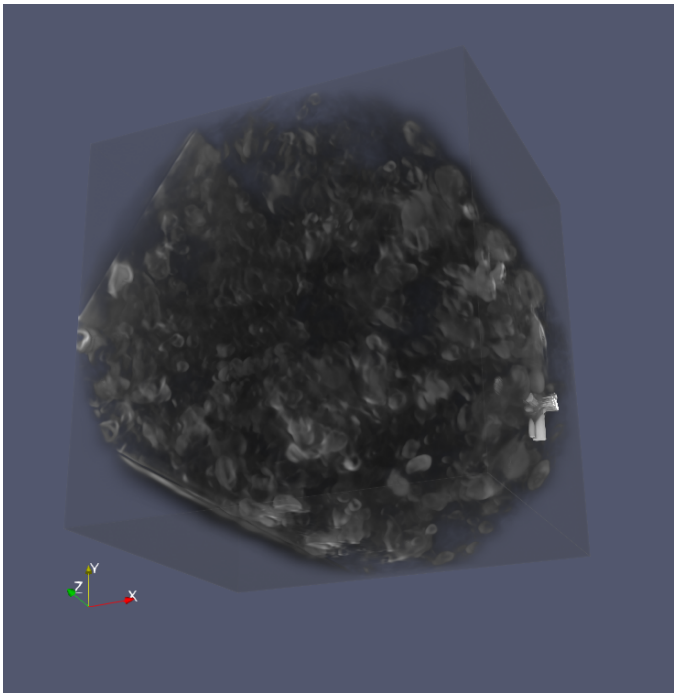

Fig. 4: Render of the deformation map for a 2560x2560x2160 volume.

the Advanced Light Source, a DOE Office of Science User Facility. We acknowledge the support of the NASA Asteroid Threat Assessment Project (Dr. E. Stern) for the characterization of meteorite materials.

References

[1] Andy Vanaerschot, Francesco Panerai, Alan Cassell, Stepan Lomov, Dirk Vandepitte, and Nagi Mansour. Stochastic characterisation methodology for 3-d textiles based on micro-tomography. 173:44–52, 08 2017.

[2] J. Michael L. MacNeil, Daniela M. Ushizima, Francesco Panerai, Nagi N. Mansour, Harold S. Barnard, and Dilworth Y. Parkinson. Interactive volumetric segmentation for textile micro-tomography data using wavelets and nonlocal means. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(4):338–353, 2019.

[3] H A Bale, A Haboub, et al. Real-time quantitative imaging of failure events in materials under load at temperatures above 1,600C. *Nat Mater*, 12:40–46, 2012.

[4] Daniela M. Ushizima, Hrishikesh A. Bale, E. Wes Bethel, Peter Ercius, Brett A. Helms, Harinarayan Krishnan, Lea T. Grinberg, Maciej Haranczyk, Alastair A. Macdowell, Katarzyna Odziomek, Dilworth Y. Parkinson, Talita Perciano, Robert O. Ritchie, and Chao Yang. Ideal: Images across domains, experiments, algorithms and learning. *The Journal of The Minerals, Metals & Materials Society*, pages 1–10, Sep 2016.

[5] P. J. M. Monteiro, C. Y. Pichot, and K. Belkebir. Computer tomography of reinforced concrete. In *Materials Science of Concrete*, volume 5, pages 537–572. American Ceramics Society, 1998.

[6] Ante Buljac, Clément Jailin, Arturo Mendoza, Jan Neggers, Thibault Taillandier-Thomas, Amine Bouterf, Benjamin Smaniotto, François Hild, and Stephane Roux. Digital volume correlation: Review of progress and challenges. *Experimental Mechanics*, 58, 06 2018.

[7] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, Dec 2004.

[8] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[9] Maryana Alegro, Edson Amaro-Jr, Burlen Loring, Helmut Heinsen, Eduardo Alho, Lilla Zollei, Daniela Ushizima, and Lea T Grinberg. Multimodal whole brain registration: Mri and high resolution histology. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 194–202, 2016.

[10] Tal Y, Rubino V, Rosakis AJ, and Lapusta N. Enhanced digital image correlation analysis of ruptures with enforced traction continuity conditions across interfaces. *Applied Sciences*, 9(8):1625–1642, 2019.

[11] M. Gates, J. Lambros, and M. T. Heath. Towards high performance digital volume correlation. *Experimental Mechanics*, 51(4):491–507, Apr 2011.

[12] Mark Gates, Michael T Heath, and John Lambros. High-performance hybrid cpu and gpu parallel algorithm for digital volume correlation. *The International Journal of High Performance Computing Applications*, 29(1):92–106, 2015.

[13] Nico P. van Dijk, Dan Wu, Cecilia Persson, and Per Isaksson. A global digital volume correlation algorithm based on higher-order finite elements: Implementation and evaluation. *International Journal of Solids and Structures*, 168:211 – 227, 2019.

[14] Stéphane Roux, François Hild, Philippe Viot, and Dominique Bernard. Three-dimensional image correlation from x-ray computed tomography of solid foam. *Composites Part A: Applied Science and Manufacturing*, 39(8):1253 – 1265, 2008. Full-field Measurements in Composites Testing and Analysis.

[15] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.

[16] Jean Feydy, Benjamin Charlier, Francois-Xavier Vialard, and Gabriel Peyre. Optimal transport for diffeomorphic registration. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pages 291–299, Cham, 2017. Springer International Publishing.

[17] M. Schmitz, M. Heitz, N. Bonneel, F. Ngolè, D. Coeurjolly, M. Cuturi, G. Peyré, and J. Starck. Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM Journal on Imaging Sciences*, 11(1):643–678, 2018.

[18] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 34(4):43–59, July 2017.

[19] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc., 2013.

[20] Dimitri P. Bertsekas and David A. Castanon. The auction algorithm for the transportation problem. *Annals of Operations Research*, 20(1):67–96, Dec 1989.

[21] Rick Chartrand, Brendt Wohlberg, Kevin Vixie, and Erik Bollt. A gradient descent solution to the monge-kantorovich problem. *Applied Mathematical Sciences*, 3(22):1071–1080, 2009.

[22] Quentin Mérigot. A multiscale approach to optimal transport. *Computer graphics forum: journal of the European Association for Computer Graphics*, 30(5):1583–1592, August 2011.

[23] Adam M Oberman and Yuanlong Ruan. An efficient linear programming method for optimal transportation. *arXiv preprint arXiv:1509.03668*, 2015.

[24] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4):66:1–66:11, July 2015.

[25] Dmitriy Morozov and Tom Peterka. Block-parallel data analysis with DIY2. In *Proceedings of the 2016 IEEE Large Data Analysis and Visualization Symposium LDAV'16*, Baltimore, MD, 2016.

[26] Gretreuer Pascal. A survey of gaussian convolution algorithms. *Image Processing On Line*, 3:286–310, 2013.

[27] B. Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, 2019.

[28] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced transport problems. *Mathematics of Computation*, 87, 07 2016.

[29] E. Haber, T. Rehman, and A. Tannenbaum. An efficient numerical method for the solution of the l_2 optimal mass transfer problem. *SIAM Journal on Scientific Computing*, 32(1):197–211, 2010.