INTERMEDIATE DEFORMABLE IMAGE REGISTRATION VIA WINDOWED CROSS-CORRELATION

Iman Aganj and *Bruce Fischl*

Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School

ABSTRACT

In population and longitudinal imaging studies that employ deformable image registration, more accurate results can be achieved by initializing deformable registration with the results of affine registration where global misalignments have been considerably reduced. Such affine registration, however, is limited to linear transformations and it cannot account for large nonlinear anatomical variations, such as those between pre- and post-operative images or across different subject anatomies. In this work, we introduce a new *intermediate* deformable image registration (IDIR) technique that recovers large deformations via windowed cross-correlation, and provide an efficient implementation based on the fast Fourier transform. We evaluate our method on 2D X-ray and 3D magnetic resonance images, demonstrating its ability to align substantial nonlinear anatomical variations within a few iterations.

Index Terms—Intermediate deformable image registration (IDIR), windowed cross-correlation, fast Fourier transform.

1. INTRODUCTION

Spatial correspondences among medical images are often needed in population and longitudinal imaging studies and are typically computed through image registration. Affine (linear) transformation is generally not sufficient to account for cross-subject anatomical variation and temporal changes in an individual anatomy, thereby making the two-step affine + deformable [1] image registration necessary in many analysis pipelines.

Standard deformable registration techniques commonly perform an optimization, during which the displacement field is updated iteratively using the very local information that the image gradient provides. In the presence of large deformations, however, the locality of this information may cause the registration to converge to the wrong local optimum and/or require many iterations with the associated increased computational burden. To alleviate these issues, the displacement field is often smoothed (regularized) [2, 3], hence corrected in each region using information from a neighborhood as large as the smoothing kernel. Furthermore, given that lower-resolution versions of the images show smaller variations (in pixels), they are pyramidally exploited either explicitly by down-sampling (coarse-to-fine registration) [4] or implicitly through contracting layers of a convolutional neural network [5-8]. Nonetheless, a transformation model neither as global as affine nor as local as most current deformable ones remains desirable, as it would provide an intermediate step for registration, which might improve the overall process.

In this work, we present a new deformable registration algorithm based on windowed cross-correlation (CC). We compute local translations by masking the two images with a sliding window function that is effectively half the size of the image (in each dimension). The large size of the neighborhood that is taken into account to compute the deformation field at each pixel makes our method a suitable intermediate deformable image registration (IDIR) approach – particularly in the presence of large deformations – to be used prior to standard (local) deformable registration. We validate our approach by applying it to the 2D X-ray and 3D magnetic resonance imaging (MRI) modalities.

In the following, we describe the proposed method in detail (Section 2) and present experimental results (Section 3) along with some concluding remarks (Section 4). Our Matlab function, `registerIDIR`, is included in our publicly available deformable image registration toolbox (www.nitrc.org/projects/msi-register).

2. METHODS

For simplicity, we start by describing our method on one-dimensional (1D) signals. Let $\tilde{I}, \tilde{J}: [\![0, N-1]\!] \to \mathbb{R}$ be two 1D discrete-domain signals of length N to be registered, where $[\![a, b]\!] \coloneqq \{a, a+1, \dots, b\}$ for $a, b \in \mathbb{Z}$. We would like to compute the transformation $T: [\![0, N-1]\!] \to \mathbb{R}$, or equivalently the displacement field $u_n \coloneqq T_n - n$, which makes $\tilde{J} \circ T$ similar to $\tilde{I}$.

2.1. Global Cross-Correlation

To avoid wrap-around effects during circular CC (see below), we first zero-pad $\tilde{I}$ and $\tilde{J}$ on their right sides, resulting in I and J of length $2N-1$, which we consider periodic outside their domain (i.e., $I_n = I_{n \bmod 2N-1}$). We then use the fast Fourier transform (FFT) to divide the frequency components of each signal by their magnitudes, to benefit from the more effective reflection of the degree of alignment by CC when the image contains only the *phase* information [9, 10].

To find the global translation that best aligns two images, the following global CC, $C: [\![-(N-1), N-1]\!] \to \mathbb{R}$, has been traditionally used [9-11]:

$$C_k \coloneqq (I \star J)_k = \sum_{n=0}^{2N-2} I_n J_{n+k}. \quad (1)$$

The sum is equivalent to a circular convolution that can be efficiently computed in the frequency domain via FFT in $\mathcal{O}(N \log N)$, as $\mathcal{F}^{-1}\{\mathcal{F}\{I\}^*.\mathcal{F}\{J\}\}$. The shift $\hat{k}$ maximizing C_k is then regarded as the translation best aligning the two signals. This global translation-only registration, thus, simply produces the constant displacement field $u_n = \hat{k}$. Next, we will extend the CC so it can produce a spatially varying 1D deformation field.

2.2. Windowed Cross-Correlation

As described above, global CC takes the entire domain of the two images into account to produce a single optimal global translation. A deformation field, on the other hand, is a set of *local* translations

Author Manuscript – *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Cartagena de Indias, Colombia, 2023.

https://doi.org/10.1109/ISBI53787.2023.10230715

Table 1. Components of $(w_\delta I \star w_\delta J)_k$.

i	$A_{i,k}$	$B_{i,\delta}$
1	$(\cos(\alpha n)\, I_n \star \cos(\alpha n)\, J_n)_k$	1
2	$(\sin(\alpha n)\, I_n \star \cos(\alpha n)\, J_n)_k + (\cos(\alpha n)\, I_n \star \sin(\alpha n)\, J_n)_k$	$\sin(\alpha\delta)\cos(\alpha\delta)$
3	$(\sin(\alpha n)\, I_n \star \sin(\alpha n)\, J_n)_k - (\cos(\alpha n)\, I_n \star \cos(\alpha n)\, J_n)_k$	$\sin^2(\alpha\delta)$

that are optimized while focusing on smaller regions in the image. CC can still be used for this purpose if the images are masked with a window function centered at the desired region, so the displacement (i.e., local translation) is computed with a focus on the image contents of that region (along the same lines as how local correlation coefficient has been exploited [12]). To compute the displacement u_δ for a region centered at $\delta \in [\![0, N-1]\!]$, we propose the following smooth window function, $w_\delta : [\![0, 2N-2]\!] \to \mathbb{R}$:

$$w_{\delta,n} \coloneqq \cos\big(\alpha(n-\delta)\big) \quad , \quad \alpha \coloneqq \frac{\pi}{2N-1}. \tag{2}$$

The window function is chosen in such a way that, when multiplied by the zero-padded image, for any $\delta \in [\![0, N-1]\!]$, the image contents (at $n \in [\![0, N-1]\!]$) only coincide with the central (positive) lobe of the window; given that $\alpha(n-\delta) \in (-\pi/2\,, \pi/2)$. The deformation field is then computed as:

$$u_\delta = \operatorname*{argmax}_k C_{\delta,k} \coloneqq \operatorname*{argmax}_{k \in [\![-(N-1), N-1]\!]} (w_\delta I \star w_\delta J)_k. \tag{3}$$

This update can be used in an iterative process until convergence. To that end, we keep an overall transformation T^o (initialized as identity) and update it as $T^o \leftarrow T^o \circ T$ (recall $T_\delta = \delta + u_\delta$), and repeat the next iteration with the updated image $\tilde{J} \circ T^o$.

One can yet immediately notice a severe practical drawback for the deformation update presented in Eq. (3). Even when employing efficient FFT for CC, computing the deformation field for all N values of δ would take $\mathcal{O}(N^2 \log N)$ operations, which is prohibitively large for images. In the next section, we will show how to reduce the computational complexity, thereby making the proposed IDIR feasible.

2.3. Computational Complexity Reduction

To reduce the computational cost of the deformation update, we exploit the expansion of the cosine function and rewrite Eq. (2) as:

$$w_{\delta,n} = \cos(\alpha n)\cos(\alpha\delta) + \sin(\alpha n)\sin(\alpha\delta). \tag{4}$$

The term $(w_\delta I \star w_\delta J)$ now expands to four CCs, which can be simplified into three terms,

$$(w_\delta I \star w_\delta J)_k = \sum_{i=1}^{3} A_{i,k} B_{i,\delta}, \tag{5}$$

where $A_{i,k}$ and $B_{i,\delta}$ are derived in Table 1.

We now approximate the argmax operation in Eq. (3) by raising $(w_\delta I \star w_\delta J)$ to the power of $P \in \mathbb{N}$ to amplify and sharpen its peak,

$$\psi_{\delta,k} \coloneqq (w_\delta I \star w_\delta J)_k^P = \left(\sum_{i=1}^{3} A_{i,k} B_{i,\delta} \right)^P, \tag{6}$$

and then approximating its peak location as its center of mass:

$$u_\delta \cong \frac{\sum_{k=-(N-1)}^{N-1} k \psi_{\delta,k}}{\sum_{k=-(N-1)}^{N-1} \psi_{\delta,k}}. \tag{7}$$

Since CC of phase images can occasionally contain small negative values, P needs to be an *odd* natural number, so such negative values do not change sign (we used $P = 3$ in Section 3).

Multinomial expansion of Eq. (6) results in $M_P \coloneqq \binom{P+2}{P}$ terms (e.g., $M_3 = 10$),

$$\psi_{\delta,k} = \sum_{j=1}^{M_P} \hat{A}_{j,k} \hat{B}_{j,\delta}, \tag{8}$$

where $\hat{A}_j$ (resp. $\hat{B}_j$) is computed following the multinomial theorem, i.e., as the product of A_is (resp. B_is) for P selections of i with repetitions allowed. Following Eqs. (7) and (8), the field becomes:

$$u_\delta \cong \frac{\sum_{j=1}^{M_P} \hat{B}_{j,\delta} \sum_{k=-(N-1)}^{N-1} k \hat{A}_{j,k}}{\sum_{j=1}^{M_P} \hat{B}_{j,\delta} \sum_{k=-(N-1)}^{N-1} \hat{A}_{j,k}}. \tag{9}$$

Thanks to the separation of functions of δ from those of k in Eq. (9), the deformation field can now be computed for the entire image simultaneously. We compute the inner sum independently of δ, and then the outer sum for all values of δ. Since a fixed number of FFTs are sufficient to produce and store all A_is, the entire u is computed in one pass in $\mathcal{O}\big(N(M_P + 7\log N)\big)$, i.e. significantly less than the original $\mathcal{O}(N^2 \log N)$, thus making our approach practical.

Next, we will extend the proposed method to images.

2.4. Extension to Higher Dimensions

For a D-dimensional image of size $N_1 \times \ldots \times N_D$, with $N = N_1 \ldots N_D$ pixels, we generalize the window function in Eq. (2) as:

$$w_{\vec{\delta},\vec{n}} \coloneqq \prod_{d=1}^{D} \cos\big(\alpha_d(n_d - \delta_d)\big), \quad \alpha_d \coloneqq \frac{\pi}{2N_d - 1}. \tag{10}$$

This generates 3^D terms in Eq. (5) and $M_{P,D} \coloneqq \binom{P+3^D-1}{P}$ terms in Eq. (8) (e.g., $M_{3,2} = 165$ and $M_{3,3} = 3654$). The vector field $\vec{u}_{\vec{\delta}}$ is then computed similarly to Eq. (9), in more general $\mathcal{O}\left(N\big(DM_{P,D} + (2^{D+1} + 3^D)\log N\big)\right)$.

2.5. Implementation

In our implementation, which we have made publicly available (see Section 1), the following were considered.

We apply a *compositive* scheme similar to the diffeomorphic demons framework [13] to update the deformation field.

When dividing the Fourier transform by its magnitude to create phase-only images, for more robustness, we divide it instead by the magnitude plus 0.001 times its norm, which reduces the weighting of low-magnitude frequency components.

Moreover, to increase stability, we reduce the effects of the peaks in the CC far from the origin by weighting the numerator and denominator of Eq. (9) by an isotropic Gaussian centered at $k = 0$ with a width equal to a tenth of the mean image dimension, which can be efficiently done by equivalently multiplying each $A_{i,k}$ by a Gaussian with P times the variance. We used this CC weighting only for images where the object was cropped by the borders (such as the X-ray images in Section 3.1), as we did not notice an improvement when the object did not touch the borders (such as the brain images in Section 3.2). Further regularization (e.g., smoothing [3]) of the

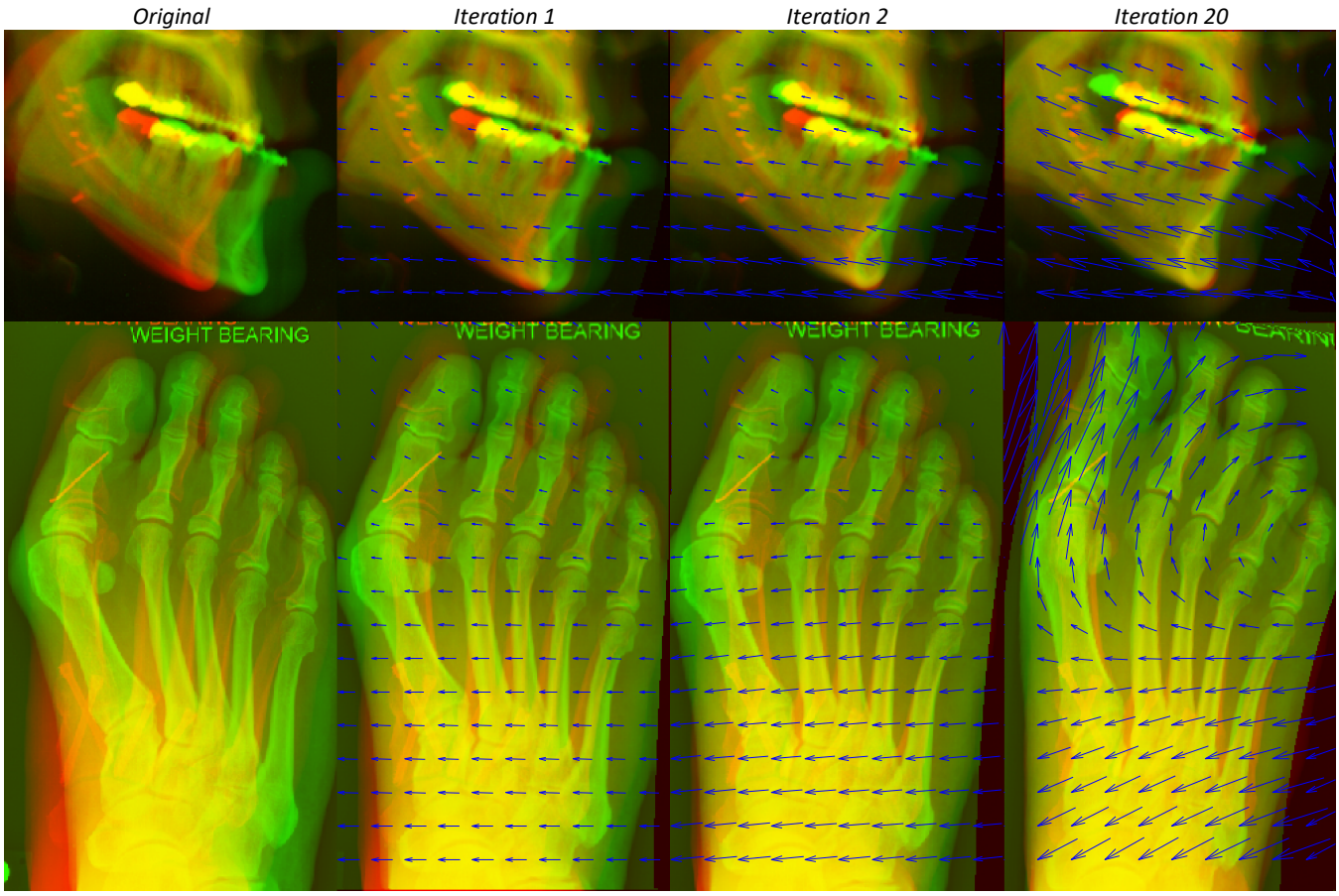

Figure 1. Different iterations of the registration of the pre- (green) to post-surgery (red) X-ray images using the proposed IDIR method.

update field is not necessary, given the inherent smoothness of the basis functions of the field (Table 1, right column).

Instead of keeping the native space of $\tilde{J}$ as the reference space, one could alternatively define forward and backward transformations, $\vec{T}_{\vec{n}}^{fw} \coloneqq \vec{n} + \vec{u}_{\vec{n}}/2$ and $\vec{T}_{\vec{n}}^{bw} \coloneqq \vec{n} - \vec{u}_{\vec{n}}/2$, and perform symmetric updates of $\tilde{J} \leftarrow \tilde{J} \circ \vec{T}^{fw}$ and $\tilde{I} \leftarrow \tilde{I} \circ \vec{T}^{bw}$ [14]. However, we empirically found the asymmetric update, which does not deal with the mid-space drift issue [15], to be more stable.

Computing the $M_{P,D}$ terms during the summation in Eq. (9) (despite no need to store them all in memory) is the most computationally expensive step of the proposed IDIR, as each $\hat{A}_j$ or $\hat{B}_j$ requires $P-1$ image multiplications. Nevertheless, by properly sorting the multinomial expansion, the results of the first $P-2$ multiplications in a term can be kept and reused for many subsequent terms, effectively reducing the number of necessary multiplications to 1 for most terms. Alternatively, these terms can be computed in parallel threads and summed.

3. RESULTS AND DISCUSSION

3.1. Experiments on 2D X-Ray Images

We applied our method first on public 2D pre- and post-operative X-ray images. In one experiment, we used a pair of jaw images of the size 170×194 acquired before and after an orthognathic surgery [16], and in another, we used a pair of foot images of the size 424×248 acquired before and after a metatarsus adductus surgery [17]. The original images, along with the results at iterations 1, 2, and 20 are shown in Figure 1. The large window in CC computation brings about the benefit of correcting large deformations within only a few iterations, but also the caveat of not fully aligning finer details that require sharp transitions in the deformation field. For comparison with standard deformable registration, we previously applied the demons method to the same jaw images, achieving accurate alignment but after about a thousand iterations [18]. Together, these results suggest the potential advantage of initializing (local) deformable registration with the proposed IDIR method.

Regarding the algorithm runtime, each iteration took 0.1 and 0.2 seconds in the first and second experiment, respectively, on a computer with a 12-core 3.6GHz (4.5GHz Turbo) Intel Xeon Gold CPU. We did not explicitly parallelize the code; however, Matlab often inherently multi-threads its internal operations.

3.2. Experiments on 3D Magnetic Resonance Images

Next, we tested our algorithm on a 3D longitudinal fetal brain MRI atlas with T2 images of the size 145×125×121, provided to the public by the Computational Radiology Laboratory of the Boston Children's Hospital [19]. We registered the first time-point of the atlas (week 21) to its last (week 38). The original images, along with iterations 1, 2, 3, and 50, are shown in Figure 2 from three viewpoints. Each iteration took 7.5 seconds on a graphics processing unit (Nvidia RTX A6000 GPU) or 40.6 seconds on the CPU (same hardware as in Section 3.1). As expected, our windowed-CC based IDIR method finds the bulk of the deformation that maps the smaller

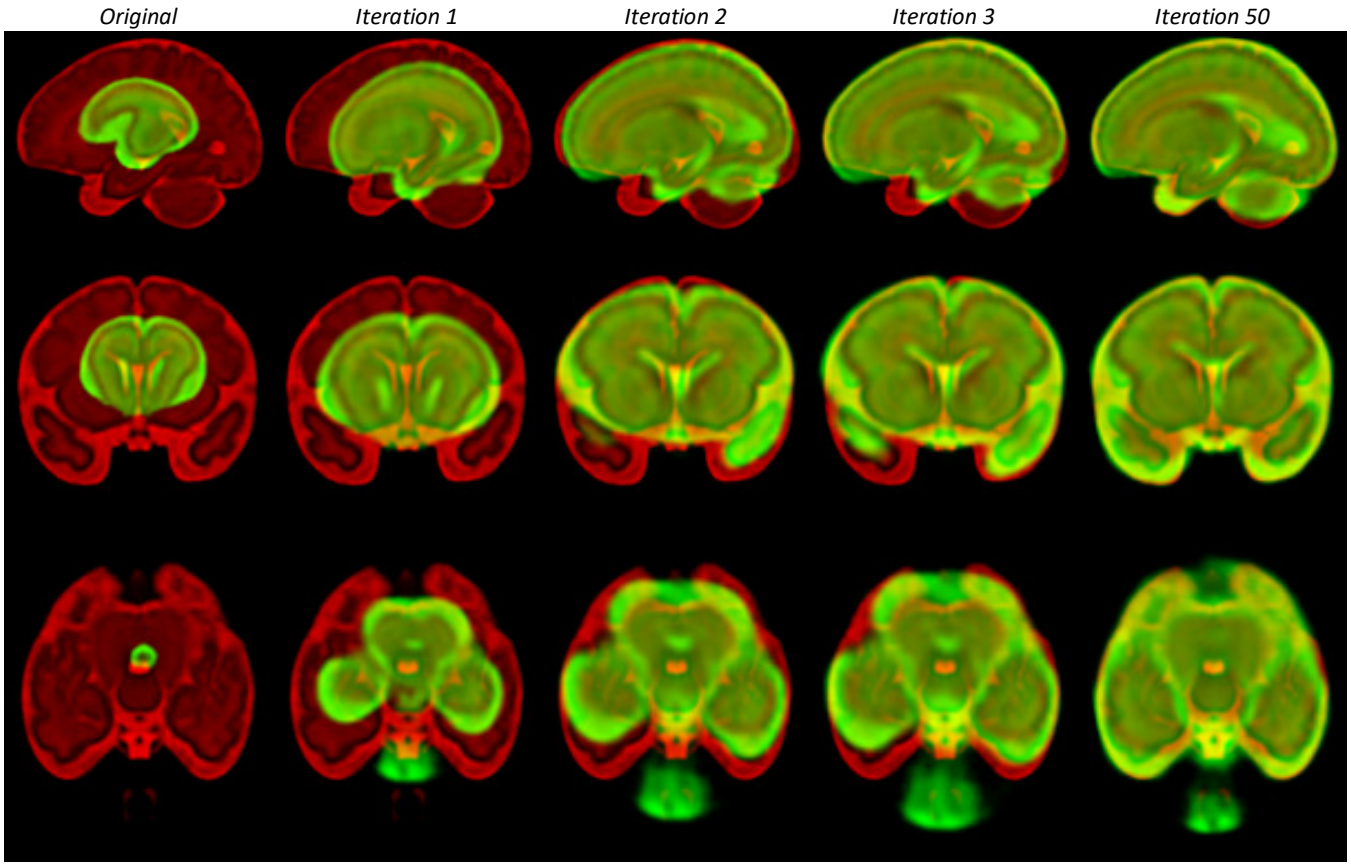

Figure 2. Central sagittal (top), coronal (middle), and axial (bottom) slices of the 3D fetal brain atlas are shown at different registration iterations (from left to right), while the first time-point (week 21, green) is registered to the last time-point (week 38, red) using the proposed IDIR method.

to the larger brain within a few iterations, eventually achieving reasonable alignment that can be perfected by standard deformable registration initialized with the IDIR result.

The dataset also contained 116 labels for different regions and structures of the brain. After propagating the labels using nearest-neighbor interpolation, we computed the portion of the voxels where the labels from the two images correctly overlapped, which is plotted for each iteration in Figure 3. The increasing nature of the plot indicates consistent improvement at each iteration, with most of the alignment occurring during the first few iterations.

For comparison, we ran the asymmetric diffeomorphic demons [13] (using our code: www.nitrc.org/projects/msi-register) on this image pair with empirically optimized (smoothing, etc.) parameters, once for 100 iterations at a single resolution, and again with a coarse-to-fine scheme for 100, 50, and 25 iterations at three resolution levels, resulting in the label overlap ratios of 0.65 and 0.76, respectively, i.e. slightly lower than the 0.77 seen in Figure 3.

When we initialized the single-level diffeomorphic demons with the IDIR result, however, the label overlap increased to 0.80.

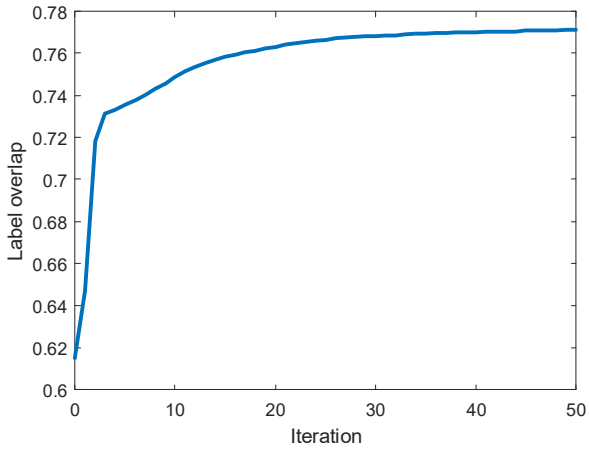

Figure 3. Label overlap ratio between the moving and reference fetal brain images at each IDIR iteration.

4. CONCLUSIONS

We have introduced a new intermediate step for deformable image registration, which exploits efficient FFT-based windowed CC computation for alignment. The large window used by the proposed IDIR method facilitates retrieval of large deformations in only a few iterations – as is apparent from our results on 2D X-ray and 3D MRI images – but may also prevent eventual capturing of fine-grained deformations. Our method is, therefore, not meant to be directly compared to standard deformable registration; instead, its potential may lie in initialization of (iterative or deep-learning based) deformable registration, offering it a head start in order to achieve a more accurate optimal solution in fewer iterations. Quantitative validation of such a two-step registration pipeline on image databases with gold-standard labels is part of our ongoing research.

5. COMPLIANCE WITH ETHICAL STANDARDS

The methodology presented in this article was evaluated retrospectively on existing publicly available de-identified images (X-ray images [16, 17] and a brain MRI atlas [19]), thereby not requiring ethical approval.

6. ACKNOWLEDGMENTS

Support for this research was provided by the National Institutes of Health (NIH), specifically the National Institute on Aging (RF1AG068261).

Additional support was provided in part by the BRAIN Initiative Cell Census Network grant U01MH117023, the National Institute for Biomedical Imaging and Bioengineering (P41EB015896, R01EB023281, R01EB006758, R21EB018907, R01EB019956, P41EB030006), the National Institute on Aging (R56AG064027, R01AG064027, R01AG008122, R01AG016495, R01AG070988), the National Institute of Mental Health (R01MH121885, RF1MH123195), the National Institute for Neurological Disorders and Stroke (R01NS0525851, R21NS072652, R01NS070963, R01NS083534, U01NS086625, U24NS10059103, R01NS105820), the NIH Blueprint for Neuroscience Research (U01MH093765), part of the multi-institutional Human Connectome Project, and the Michael J. Fox Foundation for Parkinson's Research (MJFF-021226).

Computational resources were provided through a Microsoft Azure Credit Grant by the Harvard Data Science Initiative, the Enterprise Research Infrastructure and Services at Mass General Brigham, and NIH Shared Instrumentation Grants (S10RR023401, S10RR019307, S10RR023043).

B. Fischl has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. His interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict-of-interest policies.

7. REFERENCES

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *Medical Imaging, IEEE Transactions on,* vol. 32, no. 7, pp. 1153-1190, 2013.

[2] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision,* vol. 2, no. 3, pp. 283-310, 1989.

[3] J. P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Medical Image Analysis,* vol. 2, no. 3, pp. 243-260, 1998.

[4] H. Lester, and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern Recognition,* vol. 32, no. 1, pp. 129-149, 1999.

[5] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging,* vol. 38, no. 8, pp. 1788-1800, 2019.

[6] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE Transactions on Medical Imaging,* vol. 38, no. 9, pp. 2165-2176, 2019.

[7] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis,* vol. 52, pp. 128-143, 2019.

[8] M. A. Morales, D. Izquierdo-Garcia, I. Aganj, J. Kalpathy-Cramer, B. R. Rosen, and C. Catana, "Implementation and validation of a three-dimensional cardiac motion estimation network," *Radiology: Artificial Intelligence,* vol. 1, no. 4, pp. e180080, 2019.

[9] C. Kuglin and D. Hines, "The phase correlation image alignment methed," in Proc. Int. Conference Cybernetics Society, 1975, pp. 163-165.

[10] J. J. Pearson, D. C. Hines, S. Golosman, and C. D. Kuglin, "Video-rate image correlation processor," in 21st Annual Technical Symposium, 1977, pp. 9.

[11] I. Aganj and B. Fischl, "Multi-atlas image soft segmentation via computation of the expected label value," *IEEE Transactions on Medical Imaging,* vol. 40, no. 6, pp. 1702-1710, 2021.

[12] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis,* vol. 12, no. 1, pp. 26-41, 2008.

[13] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage,* vol. 45, no. 1, Supplement 1, pp. S61-S72, 2009.

[14] P. Rogelj and S. Kovačič, "Symmetric image registration," *Medical Image Analysis,* vol. 10, no. 3, pp. 484-493, 2006.

[15] I. Aganj, J. E. Iglesias, M. Reuter, M. R. Sabuncu, and B. Fischl, "Mid-space-independent deformable image registration," *NeuroImage,* vol. 152, pp. 158-170, 2017.

[16] C. Ai. "Precautions for orthodontic treatment - Chong Ai Dental Clinic," https://lytornado.pixnet.net/blog/post/266239880-%40%E5%9C%B0%E5%8C%85%E5%A4%A9%28%E6%88%BD%E6%96%97%29%E7%89%99%E9%BD%92%E7%9F%AF%E6%AD%A3%E6%B3%A8%E6%84%8F%E4%BA%8B%E9%A0%85.

[17] "Chelsea and Westminster Hospital NHS Foundation Trust," https://www.chelwest.nhs.uk/your-visit/patient-leaflets/surgery-services/midfoot-problems.

[18] I. Aganj, M. Reuter, M. R. Sabuncu, and B. Fischl, "Avoiding symmetry-breaking spatial non-uniformity in deformable image registration via a quasi-volume-preserving constraint," *NeuroImage,* vol. 106, pp. 238-251, 2015.

[19] A. Gholipour, C. K. Rollins, C. Velasco-Annis, A. Ouaalam, A. Akhondi-Asl, O. Afacan, C. M. Ortinau, S. Clancy, C. Limperopoulos, E. Yang, J. A. Estroff, and S. K. Warfield, "A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth," *Scientific Reports,* vol. 7, no. 1, pp. 476, 2017.